DOCUMENT RESUME

ED 331 478 IR 014 963

AUTHOR	Roulin, Corentin
TITLE	Towards the European Education Metathesaurus.
INSTITUTION	EURYDICE European Unit, Brussels (Belgium).
SPONS AGENCY	Commission of the European Communities, Brussels (Belgium).
PUB DATE	90
NOTE	118p.; Translated from the French by M. Barbe and D. F. Brett. Published for the Commission; Task Force for Human Resources, Education, Training and Youth.
PUB TYPE	Reference Materials - Vocabularies/Classifications/Dictionaries (134) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Automatic Indexing; Computer System Design; *Educational Research; Foreign Countries; *Information Systems; International Cooperation; Reference Materials; *Reference Services; *Thesauri; Vocabulary
IDENTIFIERS	*EUDISED; Europe; *European Education Metathesaurus

ABSTRACT

Designed to offer a "common language" to all of the parties involved in European cooperation in education, the European Education Metathesaurus is the set made up of the European Education Thesaurus (the third edition of the EUDISED Thesaurus) and the indexing languages that will be attached to it, whether sub-thesauri, specialized vocabularies, co-thesauri, or national vocabularies. It is designed to change along with its core, the European Education Thesaurus, as it changes both to keep up with developments in education and to accommodate additional indexing languages. Current developments are the result of efforts to provide a technically consistent answer to the problems posed by its huge, changing, transdisciplinary nature while overcoming the obstacles of multilingualism and internationality. The first two chapters of this paper comprise a definition of the basic concepts, a brief description of the EUDISED Thesaurus, and a detailed presentation of the form of technical description to be used later in the paper. The remaining chapters describe the other kinds of indexing language that will be linked to the European Education Thesaurus to build the Metathesaurus set. Technical questions are given prominence with a view to providing concrete preparatory methods for the future management of the various indexing languages and their reciprocal relations, and questions relating to other conditions needed for achieving this project are also dealt with. Seven annexes include additional information on the EUDISED Thesaurus, a list of CEDEFOP (European Center for the Development of Vocational Training) Thesaurus top terms, abbreviations, a glossary, and an 11-item bibliography. (DB)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 331478

TOWARDS THE EUROPEAN EDUCATION METATHESAURUS

Eurydice

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J.A. Fernandez

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR 014 963

Published for the Commission of the European Communities
Task Force Human Resources, Education, Training and Youth by the
European Unit of EURYDICE

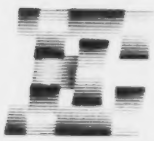
TOWARDS THE EUROPEAN EDUCATION METATHESAURUS

Corentin ROULIN
BJL Consult
for
EURYDICE European Unit
Brussels
1990

Original text in French
Translation and revision by Mrs M Barbé
Mr D F Brett

ISBN 2-87116-150-X
D/1990/4008/10

Subject to the condition that the original source is quoted, the contents of this publication may be wholly or partly reproduced except for commercial purposes.



EURYDICE European Unit
Rue Archimède 17 (Bte 17)
B - 1040 Brussels
Tel.: 02-230 03 98 Telefax: 230 65 62

PREFACE

Building Europe is a challenge involving the creation of a vast area where people, ideas and technologies can circulate freely.

Education and training play their part in this challenge, as part of the very understanding and dialogue between Europeans which is a precondition for Community cooperation aimed at establishing a new people's Europe.

Europe in recent years has seen technological advances which have led to more highly developed means of communication and a wider dissemination of knowledge. Nevertheless, these means still need to be complemented by other tools to overcome the linguistic barrier which is the major obstacle to effective exchanges. In education and training, terminology tools such as multilingual glossaries and thesauri clearly reveal the differences as well as the similarities between the various education and training systems. They therefore contribute to the promotion of information exchanges on these subjects and help in mastering knowledge of them.

Back in 1981, the Commission of the European Communities and the Council of Europe, aware of this precondition, decided to develop jointly the EUDISED multilingual thesaurus with the aim of creating the new "European Education Thesaurus". This would reflect both the wealth of educational systems and cooperation activities in this field.

This thesaurus is the result of concerted efforts by both Institutions supported by the expertise of the EURYDICE European Unit which has acted as terminology coordinator throughout.

This publication is the result of several years spent working on, and looking into, ways of pooling the educational knowledge and documentary skills possessed by the various partners involved in creating this European documentary language. Closely associated with this has been CEDEFOP, the European Centre for the Development of Vocational Training.

Given the transverse nature of education among the humanities, the Meta-thesaurus has been set up as an overall system capable of integrating the division of natural vocabulary into national terminologies and specialist vocabularies relating to specific disciplines, and the existence of documentary languages such as that used by CEDEFOP on vocational training with which closer links should be forged given the overlap of the respective areas.

This publication has, in my opinion, two basic merits. Firstly, even if this was not the author's main intention, it illustrates very clearly the proliferation of separate efforts by, and hence the differing needs of, those involved in information exchanges. Secondly, the author manages to offer a pragmatic appraisal of a complex coordinating action.

Besides the functional aim of serving the needs of those actively involved in educational documentation, this text offers a methodological approach for studying the transnational equivalents of related concepts. It thus contributes to the clarity of the very notions described in the documents for which it was only intended to serve as a storage, research or exchange tool.

I should like to stress that this publication demonstrates that building educational coope tion throughout Europe does not mean creating uniformity but rather recognising and respecting diversity. This holds true even for the implements used for handling the information. The multicultural richness underlying the variety of terminology must remain a common heritage which we must all continue to enjoy and get to know more fully.

The guidelines for development set out in this document were already agreed upon by members of the Standing Management Group of the European Education Thesaurus which met in Strasbourg in November 1988. Planning the implementation stages of the system, choosing priority issues and deciding on their implementation falls to those in charge of the European Education Thesaurus, the Council of Europe and the Commission of the European Communities. Completion of this paper coincides with publication of the new edition of the European Education Thesaurus by the Office for Official Publications of the European Communities under contract to the Task Force for Human Resources, Education, Training and Youth. It was essentially when labouring over this new edition that the ideas and intuitive feelings about future developments, and the methodological process dealt with here were brought to light.

I should therefore like to congratulate Mr C Roulin and Mrs D Hizette, responsible for terminology at the EURYDICE European Unit, together with numerous other colleagues, for the arduous and meticulous work they have done over the years with such professionalism and tenacity. Furthermore, this work has the great advantage of having been done "where it's at" rather than in a laboratory, thanks to the organisational efforts of EURYDICE and its partners in directing work on the thesaurus with a view to providing improved service for its users - those responsible for education at all levels.

I am sure that they and all specialists in comparative education will appreciate their contribution.

J A Fernández
Director of the EURYDICE European Unit

CONTENTS

Chapter 1 : Introduction

1.1. Scope 1

1.2. Indexing languages 2

1.3. The EUDISED Thesaurus 3

1.3.1. Main features 3
1.3.2. Brief historical background 4
1.3.3. Towards the European Education Thesaurus 5

Chapter 2 : Description of the EUDISED Thesaurus

2.1. Introduction 7

2.1.1. The thesaurus as a system 7
2.1.2. Description of an indexing language as a system 11
2.1.3. A metathesaurus as a system 12

2.2. Description 12

2.2.1. Aim 12
2.2.2. Elements 14
2.2.3. Relationships 17
2.2.4. Rules 24
2.2.5. Subsystems 25
2.2.6. Presentations 29
2.2.7. Use 32

Chapter 3 : Sub-thesauri and specialised vocabularies

3.1. Introduction 33

3.2. Description 34

3.2.1. Aim 34
3.2.2. Elements and relationships 37
3.2.3. Rules 37
3.2.4. Subsystems 39
3.2.5. Presentations 41
3.2.6. Use 43

3.3. Implementation 44

3.4. Specialised vocabularies 46

3.4.1. Introduction 46
3.4.2. Description and implementation 47

Chapter 4 : Co-thesauri or associated thesauri

4.1. Introduction 49

4.2. Approximation process 55

4.2.1. Conditions 55
4.2.2. Localised approach 57
4.2.3. Enlarged approach 59

Chapter 5 : National vocabularies

5.1. Introduction 61

5.2. Description 63

5.2.1. Aim 63
5.2.2. Elements 70
5.2.3. Relationships, rules and subsystems 70
5.2.4. Presentations 74
5.2.5. Use 74

5.3. Implementation 78

Chapter 6 : Towards the European Education Metathesaurus 81

Annex 1 : List of the EUDISED Thesaurus microthesauri
Annex 2 : Presentations of the EUDISED Thesaurus
Annex 3 : Instruction to the EUDISED Thesaurus user
Annex 4 : List of the CEDEFOP Thesaurus top terms
Annex 5 : Abbreviations
Annex 6 : Lexicon
Annex 7 : Bibliography

CHAPTER 1 : INTRODUCTION

1.1. SCOPE

On the occasion of its third edition, the EUDISED Thesaurus ("Multilingual Thesaurus for Information Processing in the Field of Education") is becoming the "European Education Thesaurus".

The object of this paper is to explain why its "vocation" is to become the core of the "European Education Metathesaurus ", describe what the features of such a Metathesaurus would be and secify the means that can be implemented to serve this ambitious purpose.

A thesaurus is a living thing. In its recent history we have seen that only a thorough mutation, an evolutionary process anchored in qualitative developments can enable the EUDISED Thesaurus to adapt itself to the many constraints which make up its environment.

Of course, updating at regular intervals is vital, but this will not be enough if we choose to adopt in a long-term perspective which takes into account the specificities of the field covered by the thesaurus, as well as the challenges of information exchange in tomorrow's Europe, of which the present thesaurus is a tool.

The need for this documentary device to evolve has been acknowledged previously. Today it is becoming truly possible. Indeed, lessons have been learned from managing the thesaurus, from the reflections and experiments to which it has given rise, the needs of those involved are becoming more precise, and technical means are available.

All together, these various elements have brought about this project, the "European Education Metathesaurus", the relevance and feasibility of which we now propose to demonstrate.

*

The first two chapters comprise a definition of the basic concepts, a brief description of the EUDISED Thesaurus, and a detailed presentation of the form of technical description to be used later in this paper.

The following chapters describe the other kinds of indexing language which we propose to link to the European Education Thesaurus so as to build a set that we shall call Metathesaurus.

Technical questions are given prominence with a view to providing concrete preparatory methods for the future management of the various indexing languages and their reciprocal relationships.

The questions relating to other conditions needed for achieving this project are also dealt with.

1.2. INDEXING LANGUAGES

The international standard defines an **indexing language** as *"a controlled set of terms selected from natural language and used to represent, in summary form, the subject of documents"* (ISO 5964-1985).

The **indexing** operation is part of the process for recording a document into a documentary system. By representing its content (the subject(s) dealt with) by means of several terms selected from the indexing language[1].

The user of the documentary system will draw terms from the same set to formulate a search statement [2] in the course of an **information retrieval** process, whose purpose is to retrieve the documents dealing with the subject that he is interested in.

Hence, an indexing language is a **common language** between the managers and users of a documentary system.

A **thesaurus** is defined as *"the vocabulary of a controlled indexing language, formally organized so that the a priori relationships between concepts (for example as "broader" and "narrower") are made explicit"* (id.).

It is a collection of "chosen terms" (known as "preferred terms" or **descriptors**), some with explanatory notes, (**scope notes**) which is supplemented with other terms from natural language that cannot be used in indexing (non-preferred terms or **non-descriptors** for which the user is redirected to the descriptor to be used).

---

[1] A **classification**, which also represents the content, is less flexible (however precise the code of classification may be, only one is assigned to each document) and is less readable.

Example

Title : Documentation - Guidelines for the Establishment and Development of Monolingual Thesauri

Classification : UDC 025.48

Indexation : *documentation, subject indexing, information retrieval, thesaurus, multilingual thesaurus, preparation, rules.*

[2] Other elements for recording the document in the documentary system may generally be used as retrieval criteria : retrieval on the basis of the author's name, the title, etc.; a search statement is the combination of retrieval criteria by means of the Boolean operators 'and', 'or', 'except'.

Examples

- If a user wants to identify the documents dealing with educational policy written by Mr. DELORS which were recorded into a documentary system using the EUDISED Thesaurus as indexing language, he will compose the statement

  (Author = DELORS) 'and' (subject indexing = EDUCATIONAL POLICY)

- If he wants to identify the documents dealing with educational policy in Japan, he will compose the following statement (we shall henceforth limit ourselves to searches related to indexing terms):

  EDUCATIONAL POLICY 'and' JAPAN

  which will spot the documents which were assigned both these indexing terms.

- If he wants to identify the documents dealing with free education in France and Italy, the statement

  FREE EDUCATION 'and' (FRANCE 'or' ITALY)

  will enable him to retrieve the documents indexed with the couple FREE EDUCATION and FRANCE, or the couple FREE EDUCATION and ITALY.

A thesaurus is, however, much more than a mere list of authorised or forbidden terms; it is a semantically **structured** set of concepts : relationships (either hierarchical or associative) organise the vocabulary so that it can be more easily grasped, surveyed, and used in a way that emphasizes the meaning of the terms which compose it.

We shall clarify the definition of the various types of relationships used to structure a thesaurus as soon as we have described them in more detail, on the basis of the notion of **system** (i.e. a set of elements and of relationships between these elements built for a determined purpose).

Beforehand, in order to mark out the field of indexing languages, we must add that a thesaurus is characterised by its **thematic coverage** and, in some cases, by **multilingualism**.

In effect, a thesaurus is generally built for a specific context and it should therefore cover a limited number of **areas of meaning** in detail. It will be even more efficient if the authors have managed to tailor it (and can alter it regularly) according to well-defined needs [3].

Further, a **multilingual thesaurus** is defined as *"a thesaurus containing terms selected from more than one natural language. It displays not only the interrelationships between terms, but also equivalent terms in each of the languages covered"* (Id).

Each descriptor in one language has one corresponding descriptor deemed equivalent in each of the other languages in the thesaurus [4].

However, a distinction should be made between the **concept** and its **expression** in any given language :

a) the semantic relationships that structure a thesaurus are established between the concepts that it contains, not between the terms that express them;

b) each concept in a thesaurus is expressed by one descriptor in each language ; this infers that the descriptors are considered equivalent to each other and presented accordingly, and that the semantic relationships are always strictly identical from one language version to another (since they actually link the concepts behind the words).

1.3. THE EUDISED THESAURUS

1.3.1. Main features

The EUDISED Thesaurus [5] is a documentary tool with three main features : it is an indexing language, it is multilingual, it covers the field of education.

This **indexing language** is used in a variety of contexts, notably in several European documentary systems, on either a national or an international basis.

---

[3] A documentary system managing a documentary fund specialised in botany requires an indexing language whose vocabulary relating to botany is wealthier and more accurate than that necessary for a system managing funds specialised in biology and, all the more so, in sciences. There are nonetheless encyclopaedia-oriented thesauri whose ambition is to cover all the fields of knowledge, but in little detail. These are known as "macrothesauri".

Example : the O.E.C.D. Macrothesaurus.

[4] The following terms, for example, are regarded as equivalent in the EUDISED Thesaurus :
INDUSTRIE (French), INDUSTRIE (German), INDUSTRY (English), BIOMHXANIA (Greek), INDUSTRIA (Italian), INDUSTRI (Danish), BEDRIJFSLEVEN (Dutch), INDUSTRIA (Portuguese), INDUSTRIA (Spanish).

[5] EUDISED : EUropean Documentation and Information System for Education.

It is therefore a "common language" between the managers and the users of these documentary systems. It is also the "common language" through which specialists (scientists, authorities ...) from different countries exchange information and contribute to the enrichment of the data banks managed by European bodies, such as the "EUDISED R&D (Research and Development) data base" (Council of Europe) and the "Community Data Bank in the field of Education" (Commission of the European Communities), set up by EURYDICE [6].

The 1984 version of the EUDISED Thesaurus exists in each of the nine official languages of the European Communities (versions in Finnish and Turkish are being developed).

A great deal of the quality of the thesaurus came from its genuinely **multilingual** character. Indeed, in accordance with the international standard, all languages have equal status. Each concept in the thesaurus is expressed by one descriptor in each language considered, the distinctive features and wealth of which are considered by means of non-descriptors and scope notes which vary from one language to another in content and number.

Such strict equality has also been respected in the content updating (there is one delegation per language area in the Restricted Management Group). If necessary, cooperation can be carried out on a transnational basis within one language area (example : the representative of the French-speaking language area represents France, French-speaking Belgium, Luxemburg and French-speaking Switzerland).

Last but not least, the **field covered** by the thesaurus has some particular features that must be borne in mind:

a) the boundaries between this field of social sciences and other fields are not rigid; besides, its constant evolution has produced an abundant terminology -sometimes made even more complex by quarrels between factions- which calls for some "coordination" work;

b) cultural elements play a significant part in this field, authorities are very active and, as a consequence, idiosyncracies, whether national or even regional in some countries, are very numerous, thus emphasising the usefulness of a "common language" for contributing to better knowledge of the different educational systems while respecting their diversity.

1.3.2. Brief historical background

The Council of Europe decided in 1971 to create a documentary language aimed at facilitating the exchange of information on education between member countries.

The first edition was published in 1974. It was trilingual (English, German, French) and was prepared by Mr J. VIET (Maison des Sciences de l'Homme, Paris).

From the outset the approach was :

a) transnational : specialists from different countries were called upon for each of the language groups in charge of an initial collection of terms;

b) multilingual : on the basis of the terminological material compiled by each language group, equivalences were established and the vocabulary was conceptually organised. Both operations highlighted the need for reciprocal adjustments [7].

---

[6] The education information network in the European Community.

[7] *"This multilingual approach to the field of study had the great advantage of doing justice from the outset to the particular character of each language, while avoiding the mistake, which is always a source of cultural bias, of giving priority to any one of them; it also meant that the various aspects of educational problems could be covered more accurately by reflecting the research done in several language regions"* (from the Introduction to the second edition).

The same procedures were, however, not to be respected when the following language versions were created, namely in Portuguese (Brazil, 1974), Spanish (1975) and Dutch (1977). The three of them were published in separate versions and based on the separate establishment of language equivalents.

By 1981 the General Secretaries of the Council of Europe and the Commission of the European Communities had signed a cooperation agreement on permanent development and joint funding of the thesaurus.

Such a cooperation venture led to the production of additional language versions in Greek, Italian, Danish and Portuguese in 1984, and to thorough restructuring of the vocabulary that resulted in the second edition, jointly published by the two bodies, the Council of Europe and the Commission of the European Communities.

The new edition combined the content of the 1974 vocabulary and a list of new descriptors completed in 1979; a partial and provisional edition made up of five alphabetical lists (in German, English, French, Dutch and Spanish) had been circulated in 1982.

In the 1984 edition, presentation by facets was abandoned in favour of a method distributing the vocabulary according to the microdisciplines referred to (this led to the establishment of microthesauri). Each of the nine language versions comprises three types of presentation [8] :

- structured alphabetical presentation : alphabetical list of descriptors and non-descriptors with all their relationships;

- graphic presentation of the microthesauri in the form of terminographs;

- rotated alphabetical presentation of descriptors and non-descriptors.

The production of this edition was placed under the technical and scientific responsibility of Mr. G. VAN SLYPE (Bureau Marcel Van Dijk, Brussels) and processed with the ASTUTE software[9] by I/S DataCentralen (Denmark) on behalf of the Commission of the European Communities.

1.3.3. Towards the European Education Thesaurus

Under the supervision of the terminological coordinator, the EURYDICE European Unit acting for the Task Force Human Resources, Education, Training and Youth of the Commission of the European Communities, important steps towards **updating the content** have been taken since 1986.

1986 : an addendum was produced, which compiled the proposals for new descriptors that had been put forward by the thesaurus users since 1979 (known as "Addendum 86").

1987 : a sectoral extension to the thesaurus was produced in the field of new information technologies in education (known as "NIT Extension").

1988-89 following a request for contributions circulated among the thesaurus users in January 1988, the updating process for the third edition of the thesaurus started.

---

[8] For more detail, see below, 2.2.6., page 29.

[9] ASTUTE (Automated System for Thesaurus Updating, Testing and Editing) was developed by the Directorate General "Scientific and Technical Information and Management of Information" of the Commission of the European Communities.

The main stages in the updating process were :

- combination of Addendum 86 and NIT Extension together with the demands and proposals for modifications received following the request for contributions (altogether 1,600 proposals for modifications were compiled) (summer 1988),

- compilation of both language equivalents and comments provided by delegations from the Standing Management Group (winter 1988-89),

- a proposal for selecting the additions, deletions and alterations in the structure so as to guarantee consistency in updating (spring 1989).

This proposal put forward by the terminological coordinator is presently (summer 1989) being discussed and amended by the Restricted Standing Management Group, who will allow the production of the new edition as soon as the content to be updated has been established. It is scheduled to appear at the end of 1990.

A full set of characters for each of the nine languages will also be used to enrich the formal aspect. To this end, it was decided to resort to the MICT software[10] that was considered the tool best-suited to the needs of the managers of the thesaurus, which will henceforth be known as the "European Education Thesaurus".

As the fields of education and vocational training frequently overlap and cooperation bonds exist between EURYDICE and CEDEFOP (European Centre for the Development of Vocational Training), the latter has systematically been consulted since 1986 with a view to fixing a common terminology to express the concepts common to the CEDEFOP Thesaurus and the European Education Thesaurus.

As both organisations wish to bring their standards and tools for the exchange of information as close as possible to each other, the production of a joint thesaurus , the "European Education and Vocational Training Thesaurus" is being contemplated.

---

[10] MICT (Multilingual Indexing and Classification Tool) was developed by the Office for Official Publications of the European Communities.

CHAPTER 2 : DESCRIPTION OF THE EUDISED THESAURUS

2.1. INTRODUCTION

2.1.1. The thesaurus as a system

Firstly, we shall set out to describe the EUDISED Thesaurus in its present form (second edition,1984), not only to introduce the main notions used later in this paper, but also to provide a detailed illustration of the chosen method of description.

We have already referred to the notion of system, defined as a set of elements and relationships between these elements built for a determined purpose.

At first, a monolingual thesaurus can be described, somewhat simplistically, as a set of terms (the descriptors) linked to one another by semantic relationships (whether hierarchical or associative), which is built to serve as an indexing language, to be the common language between the managers and users of a documentary system - in other words, to enable the indexing and retrieval of documents in a particular field.

We now have to explain why structuring the vocabulary is essential to reach this aim. Indeed, why can't we simply say that "a thesaurus is a collection of chosen terms that constitute a common language enabling the indexing and retrieval of documents" ?

We have pointed out that the relationships which are established between the descriptors[11] facilitate the following three operations : grasping the content of the vocabulary, surveying it and using it (i.e. indexing documents and formulating search statements).

We shall give an example for each operation.

a) grasping the content of the vocabulary

An initial and fairly cursory way to structure the vocabulary involves dividing up the descriptors into microsubjects (the notion of microthesaurus will be explained later). To grasp the overall content of the thesaurus, the user of the EUDISED Thesaurus can look at the list of these microsubjects (there are 42 of them [12]).

Should the vocabulary not be structured, the user wishing to get a more accurate idea of the thematic coverage of the thesaurus could only get his bearings from the title and indications given by the authors in the introduction or else ... read through the complete alphabetical list of descriptors.

b) surveying the vocabulary

A second, essential way of structuring involves establishing hierarchical relationships between couples of descriptors [11].

---

[11] At this first stage we are dealing with monolingual thesauri and referring to relationships between descriptors, but, for the sake of accuracy, we should say that the relationships refer to concepts, each of them being expressed by a descriptor.

[12] See Annex 1.

Figure 1. The hierarchical relationship

* Descriptor A is regarded as a narrower term of descriptor B

  if A *'is a kind of'* B

  Example : SCHOOL 'is a kind of' EDUCATIONAL INSTITUTION

  or if A *'is a part of'* B

  Example : HISTORY 'is a part of' SOCIAL SCIENCES

* If A is the narrower term of B, B is said to be the broader term of A : the relationship that links A and B is reciprocal but oriented.

  This relationship is shown by the symbols

  A 'BT' B to be read : A 'has as broader term' B

  or B 'NT' A to be read : B 'has as narrower term' A.

  (the English symbols are used in all of the EUDISED Thesaurus' language versions)

* One descriptor may be the broader term of several descriptors (it may have several narrower terms).

  Example : EDUCATIONAL INSTITUTION is the broader term of (has as narrower terms) : SCHOOL, UNIVERSITY, ...

  One descriptor may at the same time be the narrower term of one descriptor and broader term of one or several other ones.

  Example : SCHOOL has as a broader term EDUCATIONAL INSTITUTION and as narrower terms : NURSERY SCHOOL, PRIMARY SCHOOL, SECONDARY SCHOOL, ...
  SECONDARY SCHOOL may in turn have a narrower term : COMPREHENSIVE SECONDARY SCHOOL

* This example is expressed by pairs of relationships :

EDUCATIONAL INSTITUTION **NT** SCHOOL
EDUCATIONAL INSTITUTION **NT** UNIVERSITY
SCHOOL **NT** NURSERY SCHOOL
SCHOOL **NT** PRIMARY SCHOOL
SCHOOL **NT** SECONDARY SCHOOL
SECONDARY SCHOOL **NT** COMPREHENSIVE SECONDARY SCHOOL

or

SCHOOL **BT** EDUCATIONAL INSTITUTION, etc.

This relationship *"is based on degrees or levels of superordination and subordination, where the superordinate term represents a class or whole, and subordinate terms refer to its members or parts"* (ISO 2788-1986) (see Figure 1).

This type of structuring results in the building of hierarchical chains (see Figure 2, page 10) thanks to which the vocabulary can be surveyed on the basis of semantic links between terms : the user of the EUDISED Thesaurus will find all the descriptors referring to types of institutions as specific terms of the descriptor EDUCATIONAL INSTITUTION, all the descriptors referring to types of schools as specific terms of the descriptor SCHOOL, etc.

c) using the vocabulary

It can be shown that the same type of structuring (i.e. the establishment of hierarchical relationships) can facilitate both the indexing of documents and the formulation of search statements.

- indexing

In this operation it is of primary importance to be able to survey the vocabulary easily. It is advisable always to use the most specific descriptor, i.e. that which describes most accurately the content of the document to be indexed.

The most specific descriptors can be found at the lowest level in the hierarchical chains. When the indexer is looking for the descriptor to be used, he will, on the basis of the term that crossed his mind, be careful to examine the specific terms of the corresponding descriptor to check whether there are more accurate descriptors which could better represent the content of the document[13].

- retrieval

The hierarchical relationships can also be used in information retrieval to refine or enlarge the search statements that lead to unsatisfactory results, e.g. because they retrieve too many general documents dealing with a given subject, or too few precise documents, etc.

To narrow down the retrieval operation, the user can replace one of the descriptors that was first used in the search statement by a more accurate one, which can be selected among the specific descriptors of the first one.

To enlarge the search statement, he can add to the descriptor that he first used other descriptors that are generic or specific terms of it, and link them -to the first descriptor and to each another- with the operator 'or' [14].

---

[13] If an indexer dealing with a document entitled "Educational Policies in Asia" first thought that he would use the descriptors EDUCATIONAL POLICY and ASIA to represent its content and then realised that only Far East countries were involved, he should prefer the specific descriptor FAR EAST to the generic descriptor ASIA, or, if the actual content allows so, he could use some of the specific descriptors of the descriptor FAR EAST : CHINA, JAPAN,...

[14] A user wishing to find documents dealing with educational policy in Asia will first formulate the statement

EDUCATIONAL POLICY 'and' ASIA

But if he is not satisfied with the result of the research, he can enlarge the question and include the two specific descriptors of ASIA : SOUTH-EAST ASIA and FAR EAST. In so doing he will be able to extract the documents dealing with the subject with more accuracy. If documents on the subject concerning a particular Asian country are also of interest, the user could embark on a series of research stages and enlarge the statement as he goes along :

1- EDUCATIONAL POLICY 'and' ASIA

2- EDUCATIONAL POLICY 'and' (ASIA 'or' SOUTH-EAST ASIA 'or' FAR EAST)

3- EDUCATIONAL POLICY 'and' (ASIA 'or' SOUTH-EAST ASIA 'or' BANGLADESH 'or' BHUTAN 'or' ... 'or' FAR EAST 'or' CHINA 'or' KOREA 'or' ...)

Figure 2. Representation of a hierarchical chain

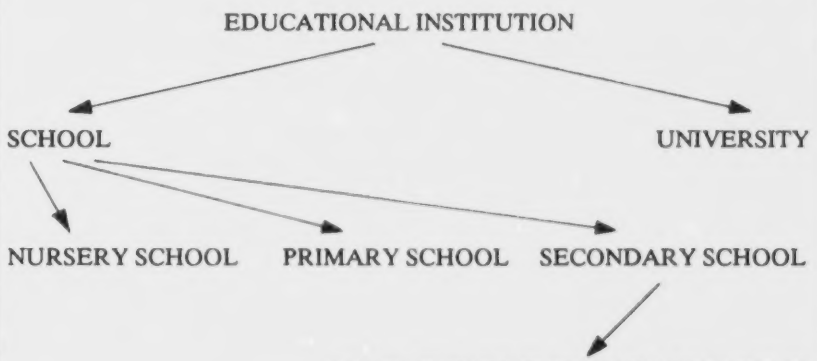
EDUCATIONAL INSTITUTION
is the broader term of / has as narrower terms

SCHOOL UNIVERSITY
is the broader term of / has as narrower terms

NURSERY SCHOOL PRIMARY SCHOOL SECONDARY SCHOOL
is the BT of / has as NT

COMPREHENSIVE SECONDARY SCHOOL

* If the hierarchical relationship is shown by a line with an arrow going from the broader term to the narrower term, we have



* If the hierarchical chain is represented by listing the narrower terms under their broader term with indentation to the right, we have

EDUCATIONAL INSTITUTION
  SCHOOL
    NURSERY SCHOOL
    PRIMARY SCHOOL
    SECONDARY SCHOOL
      COMPREHENSIVE SECONDARY SCHOOL
  UNIVERSITY

Some documentation management softwares allow automatic use of the hierarchical relationships which have been established between descriptors of the thesaurus : through posting (in indexing, for example, upward posting consists of automatically enriching human indexing with the generic descriptors of each of the descriptors chosen by the indexer [15]), or by offering the possibility of selecting easily one or several descriptors among those which are linked to a given descriptor by a hierarchical relationship (these descriptors can, for example, be suggested in a "window" appearing on the screen at the user's request).

2.1.2. Description of an indexing language as a system

To describe the EUDISED Thesaurus and the other forms of indexing languages to be used and related to each other in the framework of this project, we shall , for each, state the following points : the aim of the indexing language (what it must be used for), its structure (how it is made), the object (what its form is), and how to use it.

a) aim

Beyond the aim common to all these systems, which is to serve as indexing languages, we shall highlight their differences from two angles:

- the thematic coverage : what are the scope and features of the area of meaning that they cover within the huge field of education in Europe ?

- the possibility of communicating : a "common language" must allow an exchange of information but between whom, to what extent and how ?

b) structure

This will be described by surveying :

- the different kinds of **elements** in the thesaurus (e.g. descriptors, non-descriptors, scope notes), the method of representation (and, for existing systems, the number of these elements);

- the various kinds of possible **relationships** with the mention of the elements that they link (e.g. hierarchical relationships between descriptors);

- the **rules** that govern the relationships (e.g. a descriptor cannot be in hierarchical relationship, whether direct or indirect, with itself).

---

15 A document manually indexed with the descriptor FREE EDUCATION will be automatically given the additional descriptor EDUCATIONAL POLICY if the latter is its generic term in the thesaurus in question.

A similar system can be used to facilitate the retrieval process. In the example in note 14 above (inclusion of the specific terms of a descriptor into a search statement), the instruction given by the user would look like :

EDUCATIONAL POLICY 'and' (ASIA 'or' <all its specific terms>),
which the system will express as statement 2 in note 14.

The opposite possibility (inclusion of generic terms) would permit a statement which looks like :

EDUCATIONAL POLICY 'and' (JAPAN 'or' <its generic term>) that the system will express as

EDUCATIONAL POLICY 'and' (JAPAN 'or' FAR EAST)

that will help the user looking for documents on educational policy in Japan find those documents that consider the subject in a wider context. The document "Educational Policies in Asia" may contain information likely to be of interest to the user; it will be retrieved by

EDUCATIONAL POLICY 'and' (JAPAN 'or' FAR EAST 'or' <its generic term>) expressed by

EDUCATIONAL POLICY 'and' (JAPAN 'or' FAR EAST 'or' ASIA).

Prominence will also be given to some **subsystems**, i.e. more restricted sets of elements and relationships between these elements, which are of special interest.

c) object

We shall describe the different methods of presentation : structured, rotated, and graphic presentations while limiting ourselves to the printed form, as this is, nowadays, the most familiar to the user.

d) use

The way in which these systems should be used for document indexing and retrieval is of the utmost importance. As regards the Metathesaurus, the fact that several complementary kinds of indexing languages will be implemented at the same time make it necessary to specify how each of these tools should be used to guarantee efficiency in the exchange of information.

2.1.3. A metathesaurus as a system

The metathesaurus that we are going to describe is the system that encompasses all the other ones; its elements, i.e. the various kinds of indexing language concerned, are systems themselves.

The European Education Thesaurus is one of the elements of the "supersystem" and its very core. The other elements -the satellites of this "reference thesaurus"- are : the "sub-thesauri", which cover specialised fields relating to education, the "associated thesauri", which cover neighbouring fields, and the "national vocabularies", which cover educational situations specific to each country [16].

All these elements are therefore systems. We shall describe them not only as such, but also within the relations that they have with the "supersystem", i.e. the Metathesaurus.

While describing its elements and their dynamic interaction (relationships and rules) which structure it, we shall also show how the proposed European Education Metathesaurus is an integral part of the process to improve communication and information exchanges between the various parties involved in European cooperation in this field.

But what is really at stake besides this operational objective, is better mutual understanding, a condition necessary for cooperation between the different educational systems in tomorrow's Europe.

2.2. DESCRIPTION

2.2.1. Aim

The main function of a thesaurus, the principle of which has been described previously[17], is to allow communication; as this does not pose particular problems when the EUDISED Thesaurus is considered separately, we shall limit ourselves to the additional aspects related to the thematic coverage.

---

[16] A brief definition of these various systems can be found in the lexicon (Annex 6); they will be presented in detail in the following chapters.

[17] See above, 1.2., page 2.

The introduction to the second edition mentions that the EUDISED Thesaurus *"covers the various subjects which together constitute the interdisciplinary field of education :*

- *educational principles and systems*
- *education policy*
- *educational establishments, teachers, pupils*
- *school buildings and teaching equipment*
- *teaching curricula and subjects*
- *educational research, information on teaching*
- *psychology of education*
- *sociology of education*
- *economics of education*
- *public administration*
- *and a list of geopolitical entities".*

Delimiting "superficially" the area of meaning to be covered presents inescapable problems, and the multidisciplinary character of the field of education makes the matter even more difficult.

Nevertheless, defining "in-depth" limits is also delicate : determining the degree of specificity (accurate and restricted character) of the concepts to be retained in the thesaurus is among the questions that recur when it comes to updating the content, and it has not yet been possible to determine a rule that would be valid in all cases.

The curriculum subjects provide a good example of how bloated the vocabulary could become when both exhaustiveness and specificity are aimed for.

In this case, it was agreed to limit the coverage of the EUDISED Thesaurus to the curriculum subjects taught at primary or secondary level in at least one of the European countries.

The thesaurus' thematic cover is nonetheless also influenced by privileged contexts of use.

For example, the needs of the EURYDICE network have made it necessary to introduce vocabulary related to the administration of education and to educational structures.

In this case, the thesaurus was first limited to concepts common to all the countries and cultures involved : when a descriptor put forward for inclusion failed to have an equivalent in all the other languages (in other words, when a concept could not be expressed by a valid descriptor in a given language), it was considered that the above condition was not fulfilled and the descriptor was rejected.

This rule was made somewhat more flexible during the last updating process. Concepts absent from a given educational system or language area were introduced. In this case, the corresponding descriptor in the language version concerned can be "borrowed" from another language version [18].

This practice has been encouraged with a view to improving mutual knowledge of the various educational situations.

We shall suggest establishing links between the "national vocabularies" (which, by definition, cover situations peculiar to each country) and the "common language", embodied at European level by the thesaurus, with the same view in mind, and because this offers an additional possibility of communication.

---

[18] Some examples were already to be found in the second edition : BACHELORS DEGREE was "borrowed" from English and used as equivalent in German, Italian and Danish.

Another privileged context of use is the EUDISED R&D data base (Council of Europe), which requires a detailed vocabulary in the field of educational research.

The aim here is to meet the different specialists'needs and to make the thesaurus an efficient tool in specialised contexts.

The extension to include the vocabulary of the new information technologies in education (NIT Extension) was brought about by a Community programme developed in the field and carried out in a similar spirit.

When it was realised that the needs in a given field would exceed the thesaurus' absorption capacity (as too much vocabulary would have been introduced thus causing a great imbalance in the average degree of specificity of the thematic coverage), the idea to resort to "sub-thesauri" was launched, and some work has been done in the fields of language teaching and special education.

The ambition of this project is to provide a consistent and functional link between the "sub-thesauri" to the European Education Thesaurus, which would be their common reference thesaurus.

2.2.2. Elements

We distinguish five types of elements in the EUDISED Thesaurus.

The first three types -descriptors, non-descriptors and scope notes- are the basic elements.

Both the others -microthesauri and language versions- can be viewed as either elements or subsystems, that is to say specific "divisions" of the general system, whose practical importance is such that they can be explicitly identified.

a) descriptors (D)

Descriptors (or "preferred terms") represent concepts; they are the only elements in the thesaurus to be used in indexing.

They may occur in the form of single terms (one word only : SCHOOL) or compound terms (several words : EDUCATIONAL INSTITUTION).

Under the ASTUTE software, descriptors are represented by a chain of non-accented upper case letters[19] with a maximum of 35 positions (spaces between words included).

Each language version has 2,479 descriptors; they number (2,479 x 9 =) 22,311 in total.

b) non-descriptors (ND)

Non-descriptors (or "non-preferred terms") are synonyms or quasi- synonyms (i.e. terms that refer to very close concepts) of descriptors; they are in the thesaurus in order to strengthen the ties between the natural-language vocabulary used in documents and questions and the artificial-language vocabulary, which the indexing language is[20].

---

[19] In the Greek version characters specific to the Greek alphabet are used whereas combinations of Latin characters are used to represent typographical particularities in other versions (e.g. ü is spelled 'UE' in German, etc.).

[20] The bond with natural language first and foremost results from the choice of descriptors : it is advisable to choose words and phrases frequently used in the text of the document (this is a distinctive feature of content representation by indexing rather than classification -see note 1). On the artificial character of the indexing language, see below, 2.2.3.d), e) and f), page 19 and sqq.

Figure 3. Language indicators and number of non-descriptors and scope notes per language version

		Non-descriptors	Scope notes
D/	German	264	136
E/	English	409	136
K/	Danish	154 (-)	113 (-)
S/	Spanish	224	136
F/	French	315	136
G/	Greek	357	120
I/	Italian	457	165 (+)
N/	Dutch	839 (+)	136
P/	Portuguese	184	129
		---	---
	Total	3203	1207
	(Average)	(355,8)	(134,1)
	Maximum : (+)	Minimum : (-)	

Under ASTUTE, they are represented by the same method as that used for descriptors, and they can also be either single terms or compound terms.

The number of non-descriptors varies from one language version to another according to the lexicological wealth peculiar to each language, which differs from field to field. There is a total of 3,203 non-descriptors in the thesaurus (see Figure 3).

c) scope notes (SN)

Scope notes are short texts that either provide a definition of the terms or expressions used as descriptors or explain any usage or acceptation specific to the indexing language [21].

Under ASTUTE, they are represented by a chain of non-accented upper case letters with a maximum of 1,118 positions; a few punctuation marks are admitted.

The scope notes can also vary in number and content from one language version to another; the thesaurus has a total of 1,207 scope notes (see Figure 3).

d) microthesauri (mT)

Microthesauri have already been referred to as "microsubjects" used when first organising the vocabulary and grouping the descriptors (concepts) according to how close their meanings are, that is to say into areas of meaning.

Microthesauri can be regarded as elements since, for practical purposes, they have been identified as such by means of a two-digit serial number and a title in each language, and since they appear in the definition of the inclusive relationship.

Yet, it should be noted that microthesauri are not established *a priori* when listing the important and relevant areas of meaning but rather *a posteriori*, where the two structuring processes and their somewhat antagonistic logics meet. The first process is mainly based on the meaning and leads to the marking of coherent areas of meaning; the other one favours practical aims such as easy consultation (the creation of too many microthesauri must be avoided) or well-balanced repartition, which influences the harmony of the graphic presentation (microthesauri of comparable sizes must be created).

A more accurate description of the 42 microthesauri in the EUDISED Thesaurus[22] shows that they are rather subsystems since they form peculiar "partitions" of the general system.

Moreover, their definition calls for the notion of hierarchical chains : microthesauri can be regarded as sets in which those smaller systems known as hierarchical chains are grouped.

---

[21] Example of a scope note limited to a definition :

Descriptor : GENERAL EDUCATION
Scope note :EDUCATION WHICH, IN ITS CHOICE OF SUBJECT MATTER, DOES NOT ENVISAGE ANY KIND OF SPECIALIZATION WITH A VIEW TO PREPARING STUDENTS FOR WORK IN A PARTICULAR SECTOR

Example of a scope note which also explains the use of the descriptor :

Descriptor : REPEATING
Scope note :TO STAY IN A FORM A SECOND YEAR. DO NOT CONFUSE WITH "REPETITION", "RECITATION"

[22] See Annex 1.

e) language versions (LV)

Similarly, the nine language versions of the EUDISED Thesaurus may be regarded either as elements identified by a language indicator, i.e. a letter (see Figure 3), which appear in the definition of the omnipresent "language adherence" relationship, or as subsystems (that is to say specific "partitions" in the general system) insofar as each language version may be considered a fully-fledged thesaurus with its own elements and relationships.

2.2.3. Relationships

We distinguish seven types of relationships in the EUDISED Thesaurus.

The first three are structuring relationships. They are identical from language version to language version because they relate to concepts, even if, for the sake of convenience, we shall continue to refer to them as relationships that link the descriptors until we have provided a technical definition of the "concept". They are, namely, hierarchical, associative and inclusive relationships.

The other four share a feature with the last one in the list above : they are "compulsory" relationships, that is to say that at least one of the elements which they link cannot exist without them.

The first two relationships in this group can differ from one language version to another because they relate to non-descriptors and scope notes, which vary in number according to the language concerned. They are, namely, the semantic equivalence relationships and the "application" relationships.

The last two relationships are inherent in the multilingual character of the thesaurus. The linguistic equivalence relationship between descriptors guarantees equality and correspondence between the different language versions, whereas the "language adherence" relationship circumscribes their respective limits.

a) hierarchical relationship

This relationship was introduced above [23]; the EUDISED Thesaurus comprises 2,323 reciprocal hierarchical relationships.

Polyhierarchy is allowed - a descriptor may have more than one generic descriptor[24]. For practical purposes, this possibility has so far only been used for terms at the lowest level in hierarchical chains, i.e. terms which do not have specific descriptors.

b) associative relationship

This relationship *"covers relationships between pairs of terms which (...) are mentally associated to such an extent that the link between them should be made explicit in the thesaurus, on the grounds that it would reveal alternative terms which might be used for indexing or retrieval"* (ISO 2788-1986).

This relationship can only be established between descriptors (concepts); it is used when their natural association in common usage is not retained as a hierarchical relationship [25].

---

[23] See above, 2.1.1. b), Figures 1 and 2, page 7 and sqq.

[24] **Example :**

BIOCHEMISTRY is a specific term of both BIOLOGY and CHEMISTRY :
BIOCHEMISTRY **BT** BIOLOGY
BIOCHEMISTRY **BT** CHEMISTRY

[25] Example : EDUCATIONAL POLICY is, among other terms, related to GOVERNMENT POLICY and SYSTEM OF EDUCATION that belong to different hierarchical chains.

The usefulness of this relationship for vocabulary processing in indexing and retrieval operations is very close to that of the hierarchical relationship, and it may be described using practically the same terms [26].

The associative relationship is reciprocal : if A is related to B, B is then related to A; it is shown by A 'RT' B or B 'RT' A, to be read B 'has as a related term' A [27].

The EUDISED Thesaurus comprises 794 associative relationships which link 1,588 descriptors (concepts).

c) inclusive relationship

This relationship links descriptors (concepts) and microthesauri. It is compulsory if we consider that the microthesauri are a distribution of the vocabulary regarded in its entirety. Any descriptor (concept) belongs to at least one microthesaurus. It may belong to more than one microthesaurus in the case of some polyhierarchies [28].

There are, therefore, slightly more inclusive relationships than descriptors (concepts); this relationship is shown by the serial number(s) of the microthesaurus (microthesauri) following the wording of the descriptor [29].

This symbol appears in the structured and rotated presentations to enable the user to go immediately from the wording of a descriptor to the graphic presentation of the microthesaurus in which it is contained.

Such use of the inclusive relationship encourages a definition in terms of a relationship between descriptors (concepts) and microthesauri, but one could also look at it from the angle of the thesaurus manager or compiler and describe it as a relationship between hierarchical chains and microthesauri.

This will be the case if the definition of the microthesaurus as a subsystem is retained (we shall henceforth stick to this choice). If the microthesauri are sets of hierarchical chains, it is the inclusive relationship between hierarchical chains and microthesauri which enables the latter to be constituted.

---

[26] See above, 2.1.1. c), page 9.

[27] **Example**

EDUCATIONAL POLICY **RT** GOVERNMENT POLICY
and GOVERNMENT POLICY **RT** EDUCATIONAL POLICY

[28] **Example**

WAIFS AND STRAYS (to be changed into ABANDONED CHILD in the third edition) is a specific term of both CHILD and SOCIALLY HANDICAPPED, and appears in two microthesauri, namely number 31 (FAMILY ENVIRONMENT) and number 26 (HANDICAP).

[29] **Examples**

- EDUCATIONAL POLICY (03)
(the title of microthesaurus number 3 is PRINCIPLES OF EDUCATION)

- WAIFS AND STRAYS (26,31)

- BIOCHEMISTRY (11)
(because for both the hierarchical relationships BIOCHEMISTRY BT BIOLOGY and BIOCHEMISTRY BT CHEMISTRY, there is only one inclusive relationship in microthesaurus number 11 (SCIENCES AND TECHNOLOGY), as BIOLOGY and CHEMISTRY both appear in this microthesaurus)

The symbol of inclusion which follows the wording of any descriptor is then inferred from the hierarchical relationships and the inclusive relationship as it has been defined here above [30].

In this case, the number of relationships of inclusion and the hierarchical chains are equal, at 243.

d) semantic equivalence relationship

This is the relationship that links each non-descriptor ("non-preferred term") to the descriptor to be used ("preferred term").

It is shown by the symbol 'USE' between the non-descriptor and the descriptor or by the symbol 'UF' (used for) between the descriptor and the non-descriptor; let A be the non-descriptor and B the descriptor :

we obtain A 'USE' B , which can be read 'to express' A 'use' B
or B 'UF' A, which can be read B 'is used to express' A [31].

This relationship covers the case of "actual" synonymy as well as a series of particular cases known as "quasi-synonymy" or "conventional synonymy", which show how artificial an indexing language is, because it was built for a determined operational purpose [32].

---

[30] This was suggested in note 28 : WAIFS AND STRAYS was displayed as included in two thesauri (numbers 26 and 31) because it belongs at the same time to two hierarchical chains that are included in two different microthesauri.

See also the third example in note 29.

[31] **Example**

- STUDENT UNION **USE** STUDENT ORGANIZATION: to express STUDENT UNION, use the descriptor STUDENT ORGANIZATION

- STUDENT ORGANIZATION **UF** STUDENT UNION : the descriptor STUDENT ORGANIZATION is used to express the notion STUDENT UNION

[32] - *"Synonyms are terms whose meanings can be regarded as the same in a wide range of contexts, so that they are virtually interchangeable"* (ISO 2788-1986).

Example

ELECTRONIC COMPUTER **USE** COMPUTER

- *"Quasi-synonyms are terms whose meanings are generally regarded as different in ordinary usage, but they are treated as though they are synonyms for indexing purposes"* (Id.).

Examples

ARBITRATION **USE** DISPUTE SETTLEMENT
CONFLICT RESOLUTION **USE** DISPUTE SETTLEMENT

THERAPEUTIC EDUCATION **USE** SPECIAL EDUCATION

- Some special cases of synonymy and quasi-synonymy :

  * variant spellings

  Example

  PUPIL-TEACHER RELATION **USE** TEACHER-PUPIL RELATION

  * acronyms

  Example

  UNITED STATES OF AMERICA **USE** USA

.../...

To serve this aim, there must be as many non-descriptors as possible in order to facilitate and control (i.e. guide) the transition from natural language into indexing language.

Indeed, the larger the number of individual cases provided for (here, terms and expressions from natural language retained as non-descriptors and referring to "valid" means to express the same concepts -namely the descriptors), the fewer possibilities there will be for differences from one indexing practice to another (because the indexers have less initiative).

We shall highlight two situations with particular interest for the further development of this paper : the cases where the semantic equivalence relationship is used to express what is actually a language equivalence relationship in some cases and a hierarchical relationship in others.

- **linguistic equivalence**

  When a term in a given language is commonly used in another language, it can appear as a non- descriptor in the latter's language version of the thesaurus (whether or not this term is the language equivalent retained in the version of the thesaurus in the first language) [33].

---

[32] (cont.)

But most acronyms are developed in a scope note rather than in the form of a non-descriptor :

Example

Descriptor : OECD
Scope note : ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

* levels of speech (more or less specialised)

Examples

TRAINING GROUP USE T-GROUP
LEGASTHENIA USE DYSLEXIA

* antonymy

From two terms with opposite meanings, only one is retained as descriptor according to the notion that"*(if) there is a gradation scale between the two antonymous terms(,) when a document deals with one of the antonymous concepts, it also more or less deals with the other one*" (VAN SLYPE,G; *les langages d'indexation : conception,construction et utilisation dans les systèmes documentaires;* Les Editions d'Organisation; Paris; page 61 ; our translation).

[33] **Example**

ABSENCE USE ATTENDANCE

Let L1 and L2 be two language versions of the thesaurus;
let D1 and D2 be two descriptors regarded as language equivalents;

- **1st case** : in L2, D1 appears as a non-descriptor of D2

Example (from NTT Extension)

L1 : English version, D1 : HARDWARE
L2 : Portuguese version, D2 : EQUIPAMENTO INFORMATICO

We have, in L2 :
HARDWARE USE EQUIPAMENTO INFORMATICO

- **2nd case** : in L2, a non-descriptor of D2 is introduced, which is a term from L1 that is not D1 (but can be a non-descriptor of D1 in L1)

Example (from NTT Extension)

L1 : English version, D1 : MICROCOMPUTER
L2 : German version, D2 : MIKROCOMPUTER

We have, in L2 :
PERSONAL COMPUTER USE MIKROCOMPUTER
(and in L1 : PERSONAL COMPUTER USE MICROCOMPUTER)

This possibility has a practical advantage : the user of the version of the thesaurus in the second language will find the term among the non-descriptors in his version, with a reference to the descriptor to be used, even though he has no direct access to descriptors (and non-descriptors) in the other language versions.

This technique can be used for fields whose vocabulary developed firstly or mainly in one language (computer science vocabulary is first and foremost English) and when it is thought preferable to create an artificial term to express a concept unknown in one country or language area. The borrowed terms are then only non-descriptors (and they afford direct access to the terms which will actually be encountered in documents or questions)[34].

- **hierarchy**

"(Upward posting) *refers to a technique in which the name of a class and also the names of its members are treated as an equivalence set, the broader term then functioning as the preferred term*".

"This technique, which is sometimes used to reduce the number of terms in an indexing language, should generally be avoided. If employed, it should be applied only to terms in the fringe area of the subject field covered by the thesaurus" (ISO 2788-1986).

This brings us back to the degree of specificity of the thematic coverage[35]: whereas a thesaurus giving detailed coverage of the media has descriptors such as JOURNAL, MAGAZINE, REVIEW as specific descriptors of PERIODICAL, the EUDISED Thesaurus displays them as non-descriptors referring to one preferred term only (PERIODICAL); the semantic relationship which links these non-descriptors belongs in effect to the "specifics to generic" type.

These terms have been retained as non-descriptors because one expects to encounter them in the documents to be indexed other enough to justify explanations of each cross-reference. In other cases, it was preferred to give the rule and to mention some cross-references in a scope note by way of examples[36].

In both these cases, the different language versions remain parallel, as they share comparable non-descriptors and translated scope notes.

---

[34]See above, 2.2.1., page 12 and note 18.

To quote a classic example (ISO 5964-1985), for the French neologism VAPOCRAQUAGE which was created as an equivalent to STEAM CRACKING in English, in the French version of the thesaurus concerned, we would have :

STEAM CRACKING USE VAPOCRAQUAGE

The same standard gives the following example :

German : BREMSSTRAHLUNG

English : BRACKING RADIATION UF BREMSSTRAHLUNG

[35]See above, 2.2.1., page 12.

[36]**Example**

Descriptor : WINTER SPORT
Scope note : INCLUDES ICE SKATING, ICE HOCKEY, SKIING, ETC.

> On the other hand, some peculiarities of one country or language area, which cannot be ascribed only to the lexicological wealth of a language, can similarly be taken into account. Non-descriptors comparable to CLASSE DE MER, CLASSE DE NATURE and CLASSE DE NEIGE, which in the French version refer to the descriptor VOYAGE SCOLAIRE (SCHOOL TRAVEL) cannot be found in any of the other language versions, as such phenomena are unknown in other countries.

The semantic equivalence relationships are equal in number to the non-descriptors and can thus vary from one language version to another. They are 3,203 in total (see Figure 3, page 15).

The multi-equivalence relationship (in which one non-descriptor refers to more than one descriptor) will be introduced later [37].

e) "application" relationship

This relationship links each scope note to the descriptor to which it applies. There are as many application relationships as scope notes, i.e. 1,207 in total (see Figure 3, page 15).

It has no special symbol in the EUDISED Thesaurus. A scope note is provided under the wording of the relevant descriptor in the structured presentation[38].

This relationship is mainly of technical interest. Whereas the scope note interests the user, the relationship itself actually concerns the thesaurus compiler or manager.

We have said that the semantic equivalence relationship and the "application" relationship are both "compulsory", i.e. neither a descriptor nor a scope note can exist without being linked to a descriptor.

Conversely, we can say that the concept expressed in a given language by a descriptor cannot be correctly grasped on the sole basis of its wording, that is, without taking into account a possible scope note and the non-descriptors of which it is the preferred term. These satellite elements are fully involved in the determination of the potentially particular meaning which the terms and phrases chosen as descriptors in this artificial language may have in comparison to their meaning in natural language.

In a multilingual thesaurus the language equivalents must be added to this set of elements which make up the concept.

f) linguistic equivalence relationship

The (inter)linguistic equivalence relationship (as opposed to semantic equivalence, which is (intra)linguistic) links the descriptors expressing the same concept in the different languages [39].

The international standard imposes strict equality between the language versions of a multilingual thesaurus : one descriptor in one language must correspond to one (and only one) equivalent descriptor in each of the other languages.

---

[37] See below, 5.2.1. b), page 65.

[38] See below, 2.2.5. a), page 25.
We shall nonetheless sometimes use the symbol 'SN/' between the descriptor and its scope note later in this paper; it should be read as 'has as a scope note'.

[39] If it is considered that there is a linguistic equivalence relationship per couple of descriptors, there are 36 couples for each group of nine equivalents and, for the thesaurus as a whole, a total of (36 x 2,479 =) 89,244 reciprocal relationships of this kind.

This relationship is marked by the language indicator followed by the symbol "/" and the wording of the equivalent; all the equivalents are listed under the wording of the concerned descriptor in the structured presentation [40].

Equivalence does not come from translation. The artificial character of an indexing language, which is determined by the use for which it is intended, leads us to warn translators against improper use of this aid, as it is the result of reciprocal adjustments which may have entailed major shifts in meaning.

The international standard (ISO 5964-1985) deals extensively with difficulties that may crop up when establishing the linguistic equivalence relationship. Between exact equivalence and non-equivalence, it distinguishes three more levels of complexity or increasing difficulty : inexact equivalence, partial equivalence and single-to-multiple term equivalence.

The standard goes on to detail proposed solutions, but examining them would lead us too far astray from our target. They are : loan terms and coined terms[41], attachment of qualifying phrases and creation of multi-equivalences[42], together with some combinations of these various means (only the first one has so far been used in the EUDISED Thesaurus).

If the compulsory equivalence relationship is an inescapable source of problems for the compilers of a multilingual thesaurus, it is also very useful:

a) it guarantees equality and consistency between the different language versions which are, by and large, "equivalent" to one another : the same concepts, identically structured;

---

[40] **Example**

INDUSTRY
D/ INDUSTRIE
F/ INDUSTRIE
G/ BIOMHXANIA
I/ INDUSTRIA
K/ INDUSTRI
N/ BEDRIJFSLEVEN
P/ INDUSTRIA
S/ INDUSTRIA

[41] See notes 32 and 33.

[42] **Examples**

1) English : FUELS (motors)
French : CARBURANT

English : FUELS (heating)
French : COMBUSTIBLE

2) English : FUELS
French : CARBURANT + COMBUSTIBLE
UF CARBURANT
UF COMBUSTIBLE

3) English : HEATING
French : CHAUFFAGE

English : SOLAR ENERGY
French : ENERGIE SOLAIRE

English : SOLAR HEATING
USE HEATING + SOLAR ENERGY

b) it forms the very basis of the possibility of communicating for those involved while each partner works in whichever language he has chosen :

- in the framework of a single documentary system, and as long as the documentary management software allows, the indexing language may be either that of the indexed document or the indexer's, and the language of the search statement may be the user's with no adverse effects whatsoever on the performance of the in formation retrieval;

- in the framework of information exchanges, the interlocutor may at once "translate" any indexing or search statement into his correspondent's language or into any language which they have agreed on.

g) language adherence relationship

Each descriptor, non-descriptor and scope note belong to one language version only.

The "language adherence" relationship is mostly implicit. Each language version is published in a separate document, only the language equivalents are listed under the wording of each descriptor in the structured presentation. Their "adherence" is then marked by the symbol "X/", X being the language indicator [40].

This relationship is basically of technical interest. It particularly concerns the thesaurus compiler or manager who must produce the different versions of the thesaurus separately [43].

2.2.4. Rules

We shall not deal with this aspect at length in this part of the paper devoted to the description of the EUDISED Thesaurus in isolation. The rules applying to the relationships are mainly inherent in their definitions (mandatory or reciprocal character, type and number of the elements linked, etc.).

Defining rules will help us specify the links between the various subsystems and the European Education Thesaurus -the core of the Metathesaurus- when we describe the other forms of indexing language which constitute the Metathesaurus : methods of attachment of the sub-thesauri and national vocabularies, compatibility and methods for approximating the associated thesauri.

As we consider the rules governing the relationships in the EUDISED Thesaurus to be mainly of interest to the thesaurus compiler or manager, we shall only mention a few examples and distinguish between absolutely mandatory rules and less strict ones.

a) Example of a rule that may not be transgressed : a descriptor (concept) may not be in either direct or indirect hierarchical relationship with itself; if this were the case, surveying a hierarchical chain from top to bottom -using generic to specific relationships as a guide- might result in never-ending loops (generic-specific sequences calling for themselves).

b) Examples of less strict rules that govern the associative relationship : a descriptor (concept) is only exceptionally related

- to another descriptor in the same hierarchical chain (and, in any case, it is never related either directly or indirectly to a generic or specific descriptor (concept) nor to a "close" descriptor (concept) (i.e. one with the same generic term));

---

[43] For management purposes, it may be considered that this relationship only links the descriptors to the language versions, as it is possible to infer the "language adherence" of the non-descriptors and scope notes from the mandatory relationships which link them to a descriptor (semantic equivalence relationship on the one hand, "application" relationship on the other hand).

In this case, the "language adherence" relationships and the descriptors are equal in number, i.e. 22,311.

- to a descriptor (concept) one generic term of which is already associated to the first or any of its generic terms (e.g. descriptors whose respective "top terms" are already linked together by an associative relationship [44]).

A thesaurus management software must ensure that the rules are respected : it could reject the relationships which transgress the strict rules and emit a warning message when a less strict rule is infringed.

2.2.5. Subsystems

We distinguish four types of subsystems in the EUDISED Thesaurus.

Experience acquired in managing the thesaurus has shown how useful it was to differentiate the first two (concepts and hierarchical chains), while the other two (microthesauri and language versions) were brought to light when we listed the elements [45].

a) concepts (C)

From a technical point of view, a concept may be defined as a set of elements (linked by relationships) which help express a notion; these elements are :

- in a monolingual thesaurus ("monolingual concept") : the descriptor, any scope note and the non-descriptors that may be attached to it;

- in a multilingual thesaurus ("multilingual concept") : all the language equivalents (descriptors) with any scope note and non-descriptors.

The concept as defined above will chiefly be dealt with during the content updating; the principle of equality and the need for reciprocal adjustments in all languages have led us to perfect a specific method of presentation which enables all the information to be gathered on a single sheet of paper.

This method of presentation, which we shall call "multilingual card" (see Figure 4), is established on the basis of one of the two possible ways of identifying the concept used by the thesaurus updating management software (page number in the working document and identification number specific to the system, both of which appear in the top right-hand corner of the card) by successively resorting to the linguistic equivalence, "application" and semantic equivalence relationships.

The "multilingual card" provides additional information :

- A **definition** for each proposed addition is requested from the proposer and subsequently translated into the two working languages of the Management Group (English and French). This item of information is given to help the other parties involved find an equivalent in their languages and to provide elements which can be useful when establishing a scope note, if such a note turns out to be necessary in their language versions. The definition vanishes when the updating is completed.

---

[44] The "top term" is the descriptor (concept) at the top of a hierarchical chain; see below, 2.2.5. b), page 27.

For example, it is considered that the associative relationship between EDUCATIONAL POLICY and GOVERNMENT POLICY makes any associative relationship useless between some of their respective specific terms, such as between LIFE-LONG EDUCATION - (specific term of the former) and CULTURAL POLICY (specific term of the latter).

[45] See above, 2.2.2. d) and e), page 16.

Figure 4. Multilingual card

283

S/ PROGRAMA OCULTO

D/ HEIMLICHES CURRICULUM
E/ HIDDEN CURRICULUM
F/ EDUCATION SOUS-JACENTE
G/ KRUFO ANALUTIKO PROGRAMMA
I/ CURRICOLO IMPLICITO
K/ SKJULT LAEREPLAN
N/ VERBORGEN CURRICULUM
P/ CURRICULO OCULTO

78

-- definition ----------

S/ SERIE DE MENSAJES QUE SE TRANSMITEN COMO CONSECUENCIA DE LA FORMA DE ESCOLARIZACION, DE SU ORGANIZACION Y PRACTICAS

E/ A SERIES OF MESSAGES TRANSMITTED AS A RESULT OF THE KIND OF SCHOOLING, ITS ORGANIZATION AND PRACTICES

F/ SERIE DE MESSAGES TRANSMIS AUX ELEVES SUITE AUX MODES D'ORGANISATION DE L'ENSEIGNEMENT ET AUX PRATIQUES SCOLAIRES

-- scope notes - notes d'application ----------

S/ CONTENIDOS REALES Y NO ESTRUCTURADOS REFERIDOS ESPECIALMENTE A NORMAS, VALORES Y OPINIONES, QUE SE TRANSMITEN EN LA PRACTICA EDUCATIVA

E/ UN-STRUCTURED INFORMAL MATERIAL, CONNECTED PARTICULARLY WITH STANDARDS, VALUES AND OPINIONS, WHICH IS TRANSMITTED THROUGH TEACHING

F/ VALEURS SOUS-JACENTES VEHICULEES PAR LES PROGRAMMES, L'ORGANISATION, LES PRATIQUES, ...

I/ CURRICOLO DI FORMAZIONE CHE DERIVA DALLE CONDIZIONI DI VITA, DALLE INTERAZIONI PERSONALI, DALLA QUANTITA E QUALITA DELLE SUGGESTIONI E DEGLI STIMOLI RICEVUTI

K/ IKKE-BEVIDSTE HOLDNINGER OG VAERDIER, SOM FORMIDLES VED SIDEN AF DET OFFICIELLE UNDERVISNINGSINDHOLD

P/ CONTEUDOS REAIS E NAO ESTRUTURADOS RELATIVOS ESPECIALMENTE A NORMAS, VALORES E OPINIOES QUE SE TRANSMITEM ATRAVES DA PRATICA EDUCATIVA

-- non-descriptors - non-descripteurs ----------

D/ HEIMLICHER LEHRPLAN

-- relationships - relations ----------

Termin. 9

TT : E/ CURRICULUM
F/ PROGRAMME D'ETUDES

BT1 E/ CURRICULUM
F/ PROGRAMME D'ETUDES

- A proposal for **hierarchical** position [46] which has to be agreed on (in some cases the hierarchical environment of the concept contributes to correct grasp of its semantic content).

- Detailed **comments** received fromeach party involved may also appear.

After debates nourished by all this information, the Management Group decides whether the new concept should be accepted or rejected and agrees on any necessary adjustments.

Figure 5 is a scheme of our operational definition of the concept. The elements are symbolised by circles and the relationships by an arrow. The number of each type of element and relationship is mentioned (between brackets). Sub-system C is represented by the rectangle in bold.

The number of concepts in the EUDISED Thesaurus and the descriptors per language version both equal 2,479.

In the structured presentation of each language version they appear in the form of "monolingual concepts" supplemented by the list of language equivalents and their hierarchical and associative relationships [47].

b) hierarchical chains (HC)

The notion of hierarchical chain has already been introduced [48]; it is definitely a subsystem on the grounds that it cannot be defined without resorting to a given type of relationship (the hierarchical relationship). Besides, we shall provide a definition of this subsystem which considers as an element the subsystem "concept" as we have just defined it : a hierarchical chain is a set of concepts linked to one another by hierarchical relationships.

A top term is a descriptor (concept) that has no generic term [49]; there is one hierarchical chain per top term : each of them comprises the top term and the set of concepts which are its direct or indirect specific terms (the set can be empty [50]).

The EUDISED Thesaurus comprises 243 hierarchical chains; under ASTUTE their height may not exceed seven levels.

They are represented as such in the graphic presentation, and also appear in the structured presentation of the descriptors corresponding to top terms [51].

---

[46] "Termin." stands for "terminograph" i.e. "microthesaurus"; see below 2.2.5. c), page 29.
"TT" stands for "top term" i.e. "hierarchical chain"; see below 2.2.5. b).

[47] See Annex 2 for an example.

[48] See above 2.1.1., Figure 2, page 10.

[49] In the example in Figure 2 the top term is EDUCATIONAL INSTITUTION.

[50] A top term may be deprived of specific terms : it is then a hierarchical chain on its own.

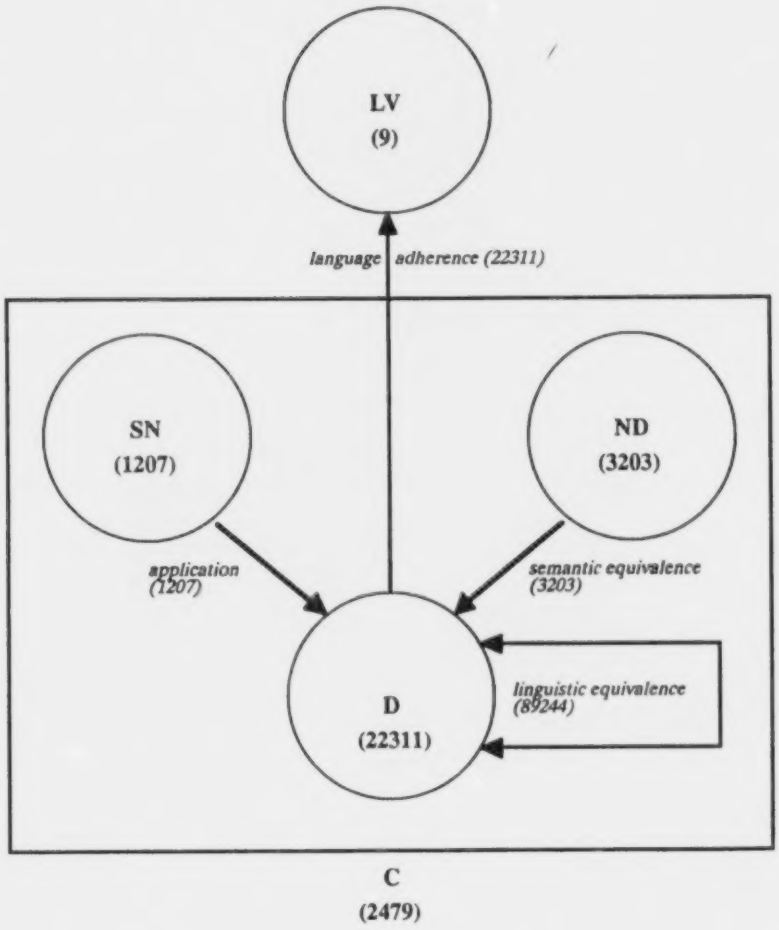
Examples

SYSTEM OF EDUCATION (04)
ENTERPRISE (32)

[51] See below 2.2.6. a) and b), page 31.

28.

Figure 5. The concept as a system



On the "multilingual card" they are identified by the symbol "TT:" followed by the descriptors, in English and in French, which correspond to the top term (see Figure 4, page 26).

c) microthesauri (mT)

In line with what we have already noted[52], and considering the subsystem "hierarchical chain" as we have just defined it, we shall define a microthesaurus as a set grouping the hierarchical chains by means of inclusive relationships.

The EUDISED Thesaurus comprises 42 of them [53]. They are identified by a serial number and a title translated into each language, which suggests giving prominence to two types of relationships whose importance is mainly technical. One may consider that :

- there is something like a linguistic equivalence relationship (a translation in this particular case) between the titles of the microthesaurus in the different language versions;

- through their serial number, the different microthesauri have a relationship which determines their succession order, which is not devoid of meaning.

The microthesauri are displayed as such in the graphic presentation, in the form of terminographs [54].

Together with the associative relationships between concepts, all the microthesauri as they have just been defined constitute the general system **thesaurus (T).**

Such a series of interlocking subsystems is displayed in Figure 6, where hierarchical and associative relationships are marked by a thinner line so as to remind the reader that they are not compulsory.

d) language versions (LV)

In line with the definitions set out above, the language versions may be considered as "partitions" or particular "views" of the general system T : each language version is a thesaurus in its own right where the concepts are "monolingual concepts" [55].

The nine language versions of the EUDISED Thesaurus are presented in separate documents, each of them comprising the three methods of presentation which we shall now introduce.

2.2.6. Presentations

We shall only briefly describe the three methods of presentation that exist in the EUDISED Thesaurus. If the reader requires a detailed description and examples, he should turn to Annex 2 in which the relevant part of the Introduction to the second edition has been reprinted.

---

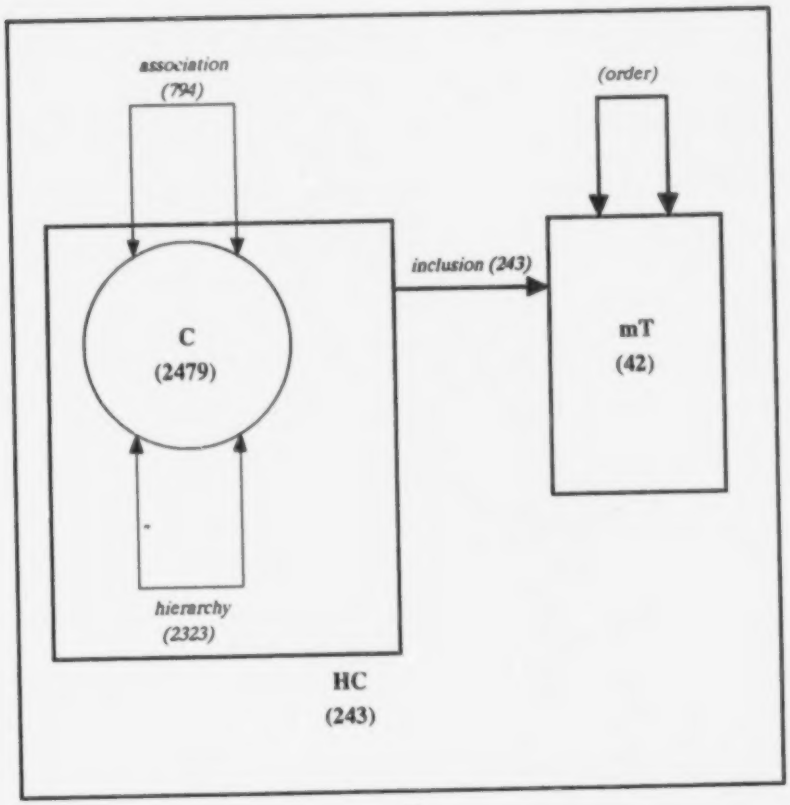
[52] See above 2.2.2. d) and 2.2.3. c), pages 16 and 18.

[53] See Annex 1.

[54] See below 2.2.6. b) page 31.

[55] See above 2.2.5.a), page 25.

Figure 6. The thesaurus as a system



a) structured alphabetical presentation

It lists all the descriptors and non-descriptors in alphabetical order.

The wording of a non-descriptor is followed by the cross-reference to the corresponding descriptor (by means of 'USE').

The wording of a descriptor is followed by :

- the serial number of the microthesaurus, the list of the eight language equivalents and any scope note (all in brackets);

- where necessary, the non-descriptors (preceded by 'UF'), the generic descriptors (preceded by 'BT1' for the directly generic terms, 'BT2' with indentation to the right for the second-degree generic terms, etc.), the specific descriptors (preceded by 'NT1', 'NT2', etc.) and the related descriptors (preceded by 'RT').

b) graphic presentation

The **terminographs** are the graphic representation of microthesauri.

Each terminograph groups the hierarchical chains which belong to the same microthesaurus on one page and provides its serial number and title.

Each hierarchical chain is displayed in the form of the third example in Figure 2, and the top term is underlined. They are written in boxes that are in turn put into a frame outside which all the descriptors which belong to other microthesauri are mentioned (the corresponding serial number is given) if they are linked by an associative relationship to a descriptor of the relevant microthesaurus.

The associative relationships are marked by bold lines linking the boxes in the case of relationships between top terms, and by thinner lines linking the descriptors themselves in the case of relationships between descriptors when neither of them are top terms, whether or not the related term is in the same microthesaurus (in which case it appears outside the frame).

Polyhierarchies are marked by a line with a double arrow.

c) rotated alphabetical presentation

This also lists all the descriptors and non-descriptors, but organises them alphabetically around each meaningful word in the wording (a compound term therefore appears as many times as there are "non-void" words[56]) and only provides the user with that minimum information which is necessary for him to refer to the other types of presentation, i.e. the thesaurus number for descriptors, the corresponding descriptor -preceded by 'USE' and followed by the serial number of the microthesaurus- for non-descriptors.

---

[56] The descriptor ADMINISTRATION OF EDUCATION appears under letter A between ADMINISTRATION and BUSINESS ADMINISTRATION and under letter E between ACCESS TO EDUCATION and ADULT EDUCATION but not under letter O (for the preposition OF).

For languages like German, Dutch or Danish, some compound words are "disagglutinated" (i.e. factored into compounds) for rotation purposes.

Example

The equivalents for REMEDIAL GYMNASTICS in German (HEILGYMNASTIK) and in Dutch (HEILGYMNASTIEK) appear under the letters H and G, and in Danish (SYGEGYMNASTIEK) under the letters S and G.

Each type of presentation is of different interest to the user :

- the structured presentation has most information and the user must refer to it to make sure that he correctly grasps the concept expressed by a descriptor, notably by consulting any scope note present;

- graphic presentation is especially practical for surveying the ways established by the associative relationships, getting an immediate, detailed view of each hierarchical chain [57] as well as of its close environment, or exploring the vocabulary on the basis of its most general organisation (i.e. the 42 microthesauri) by getting down to more and more precise levels (the 243 top terms and then each hierarchical chain in detail);

- rotated presentation affords direct access to any term that contains a specific word in order to check or find out whether a suitable descriptor exists, either in a direct way or from a non-descriptor.

2.2.7. Use

The detailed instructions addressed to the user of the EUDISED Thesaurus for indexing documents and formulating questions can be found in Annex 3, which reprints the corresponding passage from the Introduction to the second edition.

Some recommendations by the authors of the thesaurus are particularly worth reading, notably those relating to compliance, during the indexing operation, with the rules of **selectivity** (only those concepts for which the document supplies relevant information should be selected), **exhaustiveness** (no pertinent concept may be left out) and **specificity** (choosing the descriptor that most exactly corresponds to the concept which is present in the document : it is often the most precise, i.e. the most specific).

The respect of these rules influences the efficiency of the documentary system during retrieval work; efficiency can be assessed by measuring the **retrieval rate** (to what extent are all the relevant documents extracted ?) and the **precision rate** (how relevant are all the extracted documents ?).

---

[57] Each hierarchical chain appears in its entirety in the structured presentation under the entry corresponding to the top term, but it is only possible to get a partial view of it under the entries corresponding to concepts of a lower level.

Let us mention the example in Figure 2 : under the entry SECONDARY SCHOOL, the view that will be obtained of the hierarchical environment of the concept will be as follows :

SECONDARY SCHOOL

BT1 SCHOOL
    BT2 EDUCATIONAL INSTITUTION
NT1 COMPREHENSIVE SECONDARY SCHOOL

The concepts NURSERY SCHOOL and PRIMARY SCHOOL, which are located "next to it" are not mentioned, nor is, a fortiori, UNIVERSITY that is "next to" the directly generic term SCHOOL.

CHAPTER THREE : SUB-THESAURI AND SPECIALISED VOCABULARIES

3.1. INTRODUCTION

The sub-thesaurus is the first form of indexing language that we propose to graft onto the European Education Thesaurus, the core of the Metathesaurus.

Sub-thesauri are a means of turning a thesaurus into a dynamic system, an indexing language which can adapt itself with great flexibility to the evolution of the field which it covers so as to meet an increasingly wide range of needs.

The thesaurus managers intend to preserve the functional nature of the proposed tool. Sub-thesauri not only remove their deep-seated concern about keeping the thesaurus down to a manageable size and preserving the general nature of its coverage up to a certain level but also enable them to provide tailor-made answers to a potentially limitless range of needs.

Sub-thesauri allow a considerable increase in the number of users, i.e., in this case, the number of partners active in the field of education in Europe, who are able to exchange information more easily and efficiently because they use a "common language".

The attempts to develop two sub-thesauri (covering the fields of modern languages teaching [58] and special education) have not so far been successful.

Suitable data processing devices were doubtless not available, but, more importantly, in our opinion, not enough attention was paid to the design problem which should have been solved beforehand.

To close that gap we shall propose a clear and functional definition of sub-thesauri before exploring an alternative way -which is probably less costly but also less satisfactory- of resolving the problem of the coverage of specialised fields. We shall call them "specialised vocabularies" (the term "vocabularies" has been chosen to emphasise the fact that they are not structured).

The example used in the description was formulated on the basis of the "Extension to the terminology of new information technologies in the field of education" (NIT Extension , 1987), which is not a sub-thesaurus but could have become one if such a possibility had existed at the time, a fact which would have reduced the number of concepts that were included in the thesaurus proper.

For want of completed sub-thesauri, "NIT Extension" is the only example we have of completely structured specialised vocabulary to illustrate the aim of this chapter.

---

[58] See below, 3.3., page 44.

3.2. DESCRIPTION

3.2.1. Aim

a) thematic coverage

A sub-thesaurus' thematic coverage must be defined by comparison with that of the thesaurus to which it is attached and which we shall call its reference thesaurus.

A sub-thesaurus covers a more restricted field in more detail than its reference thesaurus.

Its purpose is to meet some users'need for refined indexing (e.g. documentation centres specialised in language teaching or the new information technologies in the field of education) when such needs are too large to be fully met by simply supplementing the thesaurus, or, more accurately, when the requested adjustment would cause too much vocabulary to be introduced.

Those in charge of managing the thesaurus establish the limits of its "absorption capacity" and decide whether sub-thesauri should be created. They have to take several elements into account each time that a new sub-thesaurus is considered : the nature of the need (first, the need must be acknowledged and then made quantitatively precise, in a prospective manner), the constraints imposed by the management of the necessary human, technical and financial resources and the thesaurus development policy (notably in the framework of the Metathesaurus, for which the existing or future associated thesauri[59] must be taken into account, and as regards the degree of specificity (or generality) which should be maintained in the reference thesaurus).

The sub-thesaurus is a means of breaking through the "in depth" limits of the area of meaning covered by the thesaurus; it does, however, not eradicate the problem altogether. Although creating a sub-thesaurus affords a deeper degree of specificity as regards the thematic coverage of the thesaurus and sub-thesaurus considered as a global set (in other words, the degree of specificity of the Metathesaurus' thematic coverage), the respective limitations of the thesaurus ("downwards") and sub-thesaurus ("upwards") in the new generic-specific continuum will have to be defined.

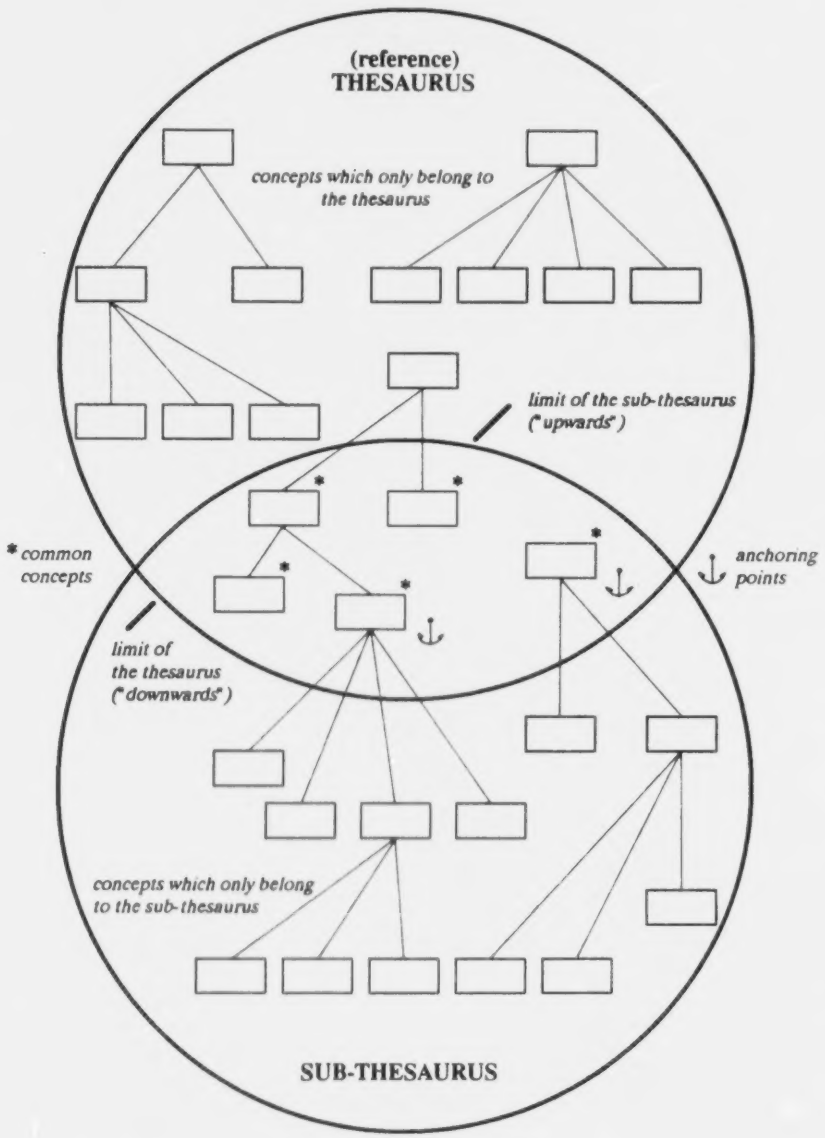
Indeed, the thesaurus and sub-thesaurus will be two non-disjointed, overlapping sets of concepts : there will be concepts that only belong either to the former or to the latter as well as concepts that are part of the intersection of both sets, since the most specific concepts in the reference thesaurus are the anchoring points of parts of hierarchical chains peculiar to the sub-thesaurus (see Figure 7) [60].

[59] See below, 4., page 49 et sqq.

[60] **Example**
(the asterisked descriptors belong to the common portion, the descriptors marked with an anchor are "anchoring points" and the descriptors in bold characters belong to the sub-thesaurus only)

COMPUTER SCIENCE*
- ARTIFICIAL INTELLIGENCE*
- COMPUTER APPLICATION* ⚓
  - **COMPUTER ASSISTED DESIGN**
  - **CONTROL TECHNOLOGY**
    - **ROBOTICS**
- COMPUTER ENGINEERING*
- PROGRAMMING*
  - PROGRAMMING ENVIRONMENT* ⚓
    - **PORTABILITY**
  - PROGRAMMING LANGUAGE* ⚓
    - **AUTHORING LANGUAGE**
    - **LOGO**
- SOFTWARE* ⚓
  - **AUTHORING SYSTEM**
  - COMPUTER GAME*
  - **COMPUTER GRAPHICS**
  - **DATA BASE MANAGEMENT SYSTEM**
  - EDUCATIONAL SOFTWARE*
  - **EXPERT SYSTEM**
  - **OPERATING SYSTEM**
  - **SPREAD SHEET**

Figure 7. Thesaurus and sub-thesaurus



So, there will be two borders to set :

- the "downward" limits of the thesaurus will be defined by applying the (mostly informal) rules that the managers have agreed to follow in other respects;

- the "upward" limits of the sub-thesaurus will be set by its compilers, who should be careful to include in the sub-thesaurus all the concepts that are useful to meet the needs expressed by future users.

b) possibilities for communication

The creation of sub-thesauri must ensure satisfactory possibilities for communication between the users and managers of :

- specialised documentary systems who use the same sub-thesaurus;

- specialised and non-specialised documentary systems (when the former use a sub-thesaurus and the latter the reference thesaurus);

- specialised documentary systems who use different sub-thesauri (all of them attached to the same reference thesaurus).

The first case poses no particular problem. A sub-thesaurus is a thesaurus in its own right, it has all the features of it and can be used alone, in which case it is the "common language" among all its users. In the other two cases, it is necessary to enable easy transition from selective, specialised indexing to indexing corresponding to the level of specificity (or, here, the level of generality) of the reference thesaurus, which remains the "common language" among all those that do not use any sub-thesaurus or use different sub-thesauri.

This leads us to round off the definition of a sub-thesaurus by imposing the possibility of automatic or semi-automatic "re-indexing" : re-indexing is the "translation" of specialised indexing into "common language" i.e. it (re-)expresses the concepts which appear in the indexing (or the search statement) by means of reference-thesaurus descriptors only.

Each descriptor used in the (specialised) original indexing is replaced by its most direct generic term in the reference thesaurus, i.e. the first descriptor in the common portion that is encountered when going up the hierarchical chain to which the original descriptor belongs [61].

Hence, a sub-thesaurus must be linked to its reference thesaurus by accurately, unambiguously defined links; this will be made possible by the constraints imposed upon the hierarchical relationships [62].

---

[61] In our example (note (60)), the indexing

EDUCATIONAL SOFTWARE, LOGO

becomes

EDUCATIONAL SOFTWARE, PROGRAMMING LANGUAGE

and the indexing

GEOGRAPHY, COMPUTER GRAPHICS

becomes

GEOGRAPHY, SOFTWARE

[62] See below, 3.2.3., page 37.

3.2.2. Elements and relationships

Like its reference thesaurus, the system "sub-thesaurus" (ST) comprises three types of basic elements : descriptors, non-descriptors and scope notes.

To provide a clear definition and facilitate subsequent technical implementation, we should distinguish, among the descriptors which appear in the sub-thesaurus, between the **sub-thesaurus descriptors (DST)** -which only belong to it- and the reference-thesaurus descriptors (D), which also belong to the sub-thesaurus (since they are in that portion common to both systems). By definition no sub-thesaurus descriptor (DST) may belong to the reference thesaurus.

The same distinction applies to non-descriptors : there are **sub-thesaurus non-descriptors (NDST),** which are attached to the sub-thesaurus descriptors, whereas the reference-thesaurus non-descriptors naturally only refer to reference-thesaurus descriptors.

The ways of representing these three elements are the same as for those in the reference thesaurus [63], and the semantic equivalence and "association" relationships have the same features.

Naturally, the linguistic equivalence and "language adherence" relationships also exist; it is possible to build sub-thesauri which include fewer language versions than the reference thesaurus (we do, however, not advise this, for reasons we shall mention later[64]).

There are also structuring relationships in the sub-thesaurus: hierarchical and associative relationships, to which the rules in force in the reference thesaurus apply, as well as some additional rules that we are now going to specify.

As for the inclusive relationship, which concerns the general organisation of the vocabulary, it will be dealt with in the section devoted to subsystems [65].

3.2.3. Rules

The rules organise the way of attaching the sub-thesaurus to its reference thesaurus.

Here the "concepts" are "multilingual concepts" in the sense which was defined when describing the EUDISED Thesaurus [66].

a) hierarchical relationship

In the sub-thesaurus a hierarchical relationship can link :

- two reference-thesaurus concepts (which belong to the common portion, i.e. the intersection of both sets);

- two concepts which only belong to the sub-thesaurus;

- one reference-thesaurus concept and one concept which only belongs to the sub-thesaurus.

---

[63] It should be noted that in the meantime the EUDISED Thesaurus will have become the "European Education Thesaurus ", and the ASTUTE software will have been replaced by the MICT management system, whose number of characters is not so limited and which uses a full set of characters in the nine languages.

[64] See below, 3.2.4. d), page 41.

[65] See below, 3.2.4. b) and c), pages 40 and 41.

[66] See above, 2.2.5. a), page 25.
See also below, 3.2.4. a), page 39.

In the last case the following rules must be respected :

- the reference-thesaurus concept belongs to the common portion *(rule 1)*;

- the reference-thesaurus concept is the broader (generic) and the sub-thesaurus concept is the narrower (specific) *(rule 2)*.

If, while compiling a sub-thesaurus, a new concept considered useful in the sub-thesaurus is to be turned, for semantic purposes, into the generic term of a concept that already exists in the reference thesaurus, the new concept must be introduced into the reference thesaurus and the latter must be adapted accordingly . Feedback is unavoidable and is also useful, as we shall explain later[67].

In addition to this, completely autonomous hierarchical chains are forbidden in the sub-thesaurus : all the concepts that are top terms of the hierarchical chains of the sub-thesaurus are reference-thesaurus concepts (in other words, they belong to the common portion) *(rule 3)*.

The consequence of this rule is that each sub-thesaurus concept is effectively linked by a hierarchical relationship, whether direct or indirect, to a reference thesaurus concept; re-indexing in the sense which we have defined is thus possible in all cases.

To these three strict rules we shall add one which we feel should be adhered to : polyhierarchies are forbidden in the sub-thesaurus *(rule 4)*; we shall justify this rule while adding certain qualifications.

We have seen that in the EUDISED Thesaurus the possibility of establishing polyhierarchies has only been used for very specific concepts (i.e. those that have no specific terms of their own).

In the portion common to the reference thesaurus and the sub-thesaurus, we feel that polyhierarchies should be forbidden for concepts which actually form anchoring points of parts of the hierarchical chains specific to the sub-thesaurus *(rule 4A)* :

- even if these concepts are the most specific ones in the reference thesaurus, they remain of a relatively generic nature when the hierarchical chains are considered as a continuum[68] (see Figure 7, page 35);

- the parts of hierarchical chains "anchored" to concepts with more than one generic term in the common portion would be reproduced in the structured and graphic presentations of the sub-thesaurus as many times as there are generic terms, and this would be altogether useless and impractical.

In the portion specific to the sub-thesaurus, we advise against polyhierarchies *(rule 4B)*, as re-indexing a sub-thesaurus concept which belongs to two different hierarchical chains would lead to "translating" that single specific concept into two relatively generic descriptors -a fact which would aggravate the loss of specificity inherent in this operation.

b) associative relationship

The sub-thesaurus comprises the associative relationships established in the reference thesaurus between the concepts which belong to the common portion, as well as associative relationships which only belong to it.

The latter can of course only be established between two concepts which only belong to the sub-thesaurus or between a sub-thesaurus concept and a common-portion concept *(rule 5)*.

---

[67] See below, 3.3., page 44.

[68] Some hierarchical chains can extend from the reference thesaurus (by their "top") to the part which only belongs to the sub-thesaurus (by their "bottom") through the common portion.

3.2.4. Subsystems

Like its reference thesaurus, the system "sub-thesaurus" may be represented by a series of the same interlocking subsystems : concepts, hierarchical chains and subsystems comparable to the microthesauri in the reference thesaurus, i.e. "micro-sub-thesauri", which we shall call "nanothesauri ".

It also includes language versions.

a) concepts

Thus, the sub-thesaurus comprises both **sub-thesaurus concepts (CST)** and reference-thesaurus concepts (C) which are in the common portion (at the top of each hierarchical chain in the sub-thesaurus, in accordance with the third rule).

This situation gives rise to two types of problem : a technical problem of "marking" the respective borders of the systems T and ST (we shall come to this when we deal with the hierarchical chains) and a problem connected to the semantic content of the reference-thesaurus concepts which serve as anchoring points to hierarchies developed in the sub-thesaurus.

Indeed, when these concepts are considered from the point of view of the reference thesaurus, we cannot take for granted that they are the most specific terms and ignore that they are in fact the generic terms of a series of sub-thesaurus concepts, which remain hidden from the user of the reference thesaurus.

We know that the hierarchical environment of a concept can contribute to grasping correctly its semantic content[69]; this holds true for those concepts, and we believe that the user of the reference thesaurus should know about their status as anchoring points in the sub-thesaurus.

That is why we recommend that in the reference thesaurus :

- the descriptors expressing these concepts be formally identified as anchoring points in the sub-thesaurus by means of some typographical device in the different presentation [60];

- the list of non-descriptors attached to them be enriched by the "upward-posting" technique [70], that is to say that some sub-thesaurus descriptors should appear as non-preferred terms of reference-thesaurus descriptors of which they are actually specific terms in the sub-thesaurus.

Such use of "upward posting" complies with the international standard quoted before : *"this technique, which is sometimes used to reduce the number of terms in an indexing language, should generally be avoided. If employed, it should be applied only to terms in the fringe area of the subject field covered by the thesaurus."* : their location in the fringe area should be clear enough to justify the development of a sub-thesaurus instead of the enrichment of the thesaurus proper.

For implementing purposes we also recommend :

- distinguishing these non-preferred terms from the actual non-descriptors by means of a special symbol in the structured presentation, so as to remind the user that the type of relationship linking these non-preferred terms to the preferred terms is halfway between the semantic equivalence relationship and the hierarchical relationship;

---

[69] See above, 2.2.5. a), page 25.

[70] See above, 2.2.3. d), page 19.

if A is the reference-thesaurus descriptor and B its specific term in the sub-thesaurus :

A 'UFS' B (Used For from Sub-thesaurus) [71]

- retaining as non-preferred terms only the sub-thesaurus descriptors that are direct specific terms or specific terms to the second degree (and rejecting their own non-descriptors) in order not to cram the reference thesaurus, as this would have the oppositie effect to that desired.

b) hierarchical chains

From a technical point of view there are no hierarchical chain specific to the sub-thesaurus, as the top of each is also part of the reference thesaurus and some chains can extend beyond it [72].

It must nonetheless be possible to split the hierarchical chains in order to edit them correctly and distribute the sub-thesaurus' hierarchical chains into the various nanothesauri; this can be done if the inclusive relationship is defined as a relationship between **concepts which are top terms** in the sub-thesaurus and **nanothesauri.**[73]

---

[71] **Example**

SOFTWARE *⚓
UF SOFTWARE TOOL
UFS AUTHORING SYSTEM
UFS COMPUTER GRAPHICS
UFS DATA BASE MANAGEMENT SYSTEM
... ...
NT COMPUTER GAME*
NT EDUCATIONAL SOFTWARE*

[72] For instance, a sub-thesaurus on language teaching might not include the generic descriptor LANGUAGE TEACHING but its specific terms only.

Indeed, it may be considered :

- either that the indexing of almost all the documents listed in a documentary system specialised in that field may include that descriptor (all the documents deal with that subject), which would then lose its discrimination capacity and hence its usefulness (retrieval concerned with that descriptor would select all the recorded documents);

- or that, for this very reason, it is never used (besides, the rule of specificity of indexing imposes the need to choose the descriptors that most closely match the concepts in the document : the latter are almost always more accurate and therefore more specific).

[73] In the reference thesaurus, the hierarchical chains end with the most specific reference-thesaurus concepts (C) : when the latter have hierarchical relationships with sub-thesaurus concepts (CST), they are not taken into account (regardless of the fact that some are retained as sources of non-preferred terms).

In the sub-thesaurus the upper limit of each hierarchical chain cannot be based on the nature of the concepts encountered when surveying it upwards (at the top only reference-thesaurus concepts (C) are met, which nothing differentiates a priori), nor can it be based on the absence of hierarchical relationship with a generic term, which characterises top terms in the reference thesaurus.

The reference-thesaurus concepts must, however, be identified by the manager and the management system.

That is why in the sub-thesaurus we define the inclusive relationship as a relationship between concepts and nanothesaurus and, more accurately, between concepts (C) which are top terms in the sub-thesaurus and nanothesaurus.

Indeed, problems could arise from the wording retained for the definition of microthesauri in the description of the EUDISED Thesaurus (see above, 2.2.3. c), page 18) : how could nanothesauri be defined as sets of hierarchical chains made up by means of the inclusive relationship if, as we have stated, there are, technically, no hierarchical chains specific to the sub-thesaurus ?

Nevertheless, there is no great practical difference :

- in the thesaurus the inclusive relationship between the hierarchical chains may be given concrete expression either by a relationship between top terms and microthesaurus (the symbol of inclusion which follows the wording of any descriptor is then deduced from the hierarchical relationships and the inclusive relationship thus established), or by a relationship between each concept (whether top term or not) and the microthesaurus to which it belongs; .../...

c) nanothesauri (nT)

They are the sub-thesaurus equivalents to the reference thesaurus microthesauri .

Nanothesauri can mark divisions into areas of meaning which differ from those marked by microthesauri. Like microthesauri, they are delineated so as to satisfy the need for both semantic coherence and balance in the vocabulary distribution [74].

The set formed by the nanothesauri and the associative relationships between sub-thesaurus concepts make up the system **sub-thesaurus** (ST); the corresponding diagram is provided in Figure 8.

d) language versions

We have already mentioned that we have reservations about the possibility of developing sub-thesauri comprising fewer language versions than their reference thesaurus.

It is true that a higher number of languages entails a more complex task and considerably higher costs , but we think that the following should nonetheless be taken into account :

- the European vocation of the Metathesaurus must be shown in each of the tools that compose it;

- using "upward posting" only in the case of the languages for which there is a version of the sub-thesaurus would cause imbalance between the language versions, and this should be avoided;

- unavoidable feedback from the development of a s ub-thesaurus to the content and structure of the reference thesaurus itself can be integrated more easily and rapidly if it is based on work in all the languages from the outset [75].

3.2.5. Presentations

The relative complexity of the set "thesaurus-sub-thesaurus", mainly due to the "dual adherence" of some of the concepts and to the fact that divisions which differ between micro- and nanothesaurus were allowed, lead us to advise against making a complete joint edition that would present the set as a whole in a single document.

The numerous new conventions which should be introduced into the methods of presentation would make the use of such an instrument very difficult indeed. Only those in charge of the management should have available at least one joint method of presentation.

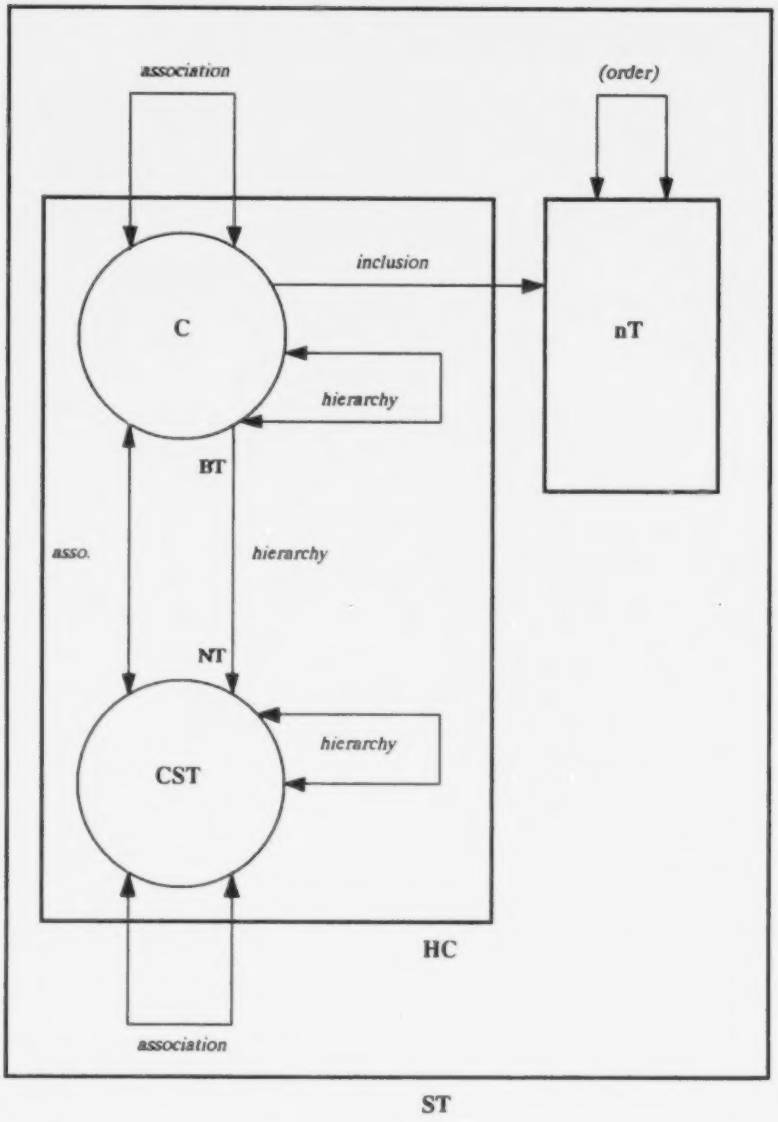
---

[73] (cont.)
- similarly, in the sub-thesaurus, the inclusive relationship may be established either between top terms in the sub-thesaurus and nanothesaurus only (this solution is schematised in Figure 8), or between each concept that belongs to the sub-thesaurus (C or CST) and the nanothesaurus to which it belongs (the determination of the upper limit of a hierarchical chain is then made by spotting the concepts C belonging to the common portion which have no generic term also belonging to the common portion).

[74] For a microthesaurus (13 LANGUAGE SCIENCES) there may be several nanothesauri that correspond, for example, to the different hierarchical chains in the reference thesaurus : LANGUAGE SCIENCES, COMMUNICATION, LANGUAGES, etc.

[75] Those responsible for the management of the Metathesaurus will then be given the assurance that they needn't check whether extra interlinguistic adjustments are necessary or not in the reference thesaurus; conversely, if proposals for feedback must be examined in detail, there would be a risk that decisions taken in a more restricted multilingual context would be contested, refused or modified at thesaurus-management level, which would cause other feedback to the sub-thesaurus this time.

Figure 8. The sub-thesaurus as a system



As for the users, they will choose either the reference thesaurus, if it is sufficient for their needs, or the sub-thesaurus, which should be designed in such a fashion that it can be used alone.

a) reference thesaurus

There are few changes from the description of the EUDISED Thesaurus :

- prominence can be given to all the descriptors corresponding to common concepts by means of some typographical device. Moreover, in each type of presentation, the anchoring points of the sub-thesaurus (sub-thesauri) are given prominence. The user's attention is thus drawn to the fact that one or several structured specialised vocabularies exist (this indication is given in all the language versions, even those for which there is no corresponding version of the sub-thesaurus, in order to guarantee information to all users and to those who might cooperate in developing sub-thesauri);

- in accordance with what we have stated, the list of the non-descriptors attached to the descriptors corresponding to these "anchoring-point concepts" is supplemented with "upward-posting" and appropriate symbols are introduced [76].

b) sub-thesaurus

This is presented in a separate document and may be republished at different intervals to its reference thesaurus.

For each language version and in line with the three forms of presentation in the reference thesaurus, it presents the common concepts and the concepts which only belong to the sub-thesaurus, and distinguishes the anchoring points with the same typographic device as that used in the reference thesaurus.

The graphic presentation respects the division into nanothesauri.

3.2.6. Use

The sub-thesaurus is thus a thesaurus in its own right which can be used alone (as long as it includes enough concepts outside the specific field covered - the "common portion" may be considerable).

Re-indexing, made possible by the constraints imposed upon the hierarchical relationships, is achieved either automatically by a computer programme or manually. It has two stages :

- each sub-thesaurus concept that appears in the indexing is re-expressed by the descriptor corresponding to its anchoring point in the reference thesaurus, i.e. the concept to which the hierarchy that it belongs to is attached;

- any doubles that result from the first stage are deleted.

This operation is necessary, for example when records comprising indexations based on the sub-thesaurus are included in a documentary system which only uses the reference thesaurus as an indexing language.

It is also possible to have a documentary system in which both the specialised indexing and its "translation" into "common language" appear : indexing is originally performed at the most specific level on the basis of the sub-thesaurus, and after re-indexing, the result is stored in a distinct field. The user chooses the field in which to search.

---

[76] See above, 3.2.4. a), page 39.

Conversely, when a specialised documentary system is interrogated by a user expressing a search statement by means of reference- thesaurus descriptors only, his key must be enlarged, automatically or otherwise, to include all the sub-thesaurus descriptors that are specific terms of the descriptors used [77].

The latter case shows that the thematic coverage of the reference thesaurus must be sufficiently specific in order not to jeopardise performance in such an operation.

3.3. IMPLEMENTATION

Obviously, the compilation of a sub-thesaurus is by definition at least as complex an undertaking as the compilation of a thesaurus.

Building on existing foundations (i.e. a multilingual, ready structured vocabulary) is an advantage (part of the work is done) and a drawback : making a sub-thesaurus will always engender doubts about the structure of the relevant parts in the reference thesaurus.

But such doubts are of great value to the reference thesaurus, which, given this opportunity, must be improved. Making sub-thesauri should not only afford the possibility of avoiding too much vocabulary, but also of reinforcing the relevance of the reference thesaurus' structure.

It may also lead to a reduction in volume as relatively specific concepts, which are on the whole rarely used, could be removed to a sub-thesaurus.

A decision could then be taken to compel the managers of the reference thesaurus to keep its volume universally constant, which would reassure those who fear that a thesaurus covering so wide a field is bound to grow for ever and become too bulky for practical use.

Working on more clearly delineated fields is both favourable -highly qualified experts may be relied on- and unfavourable. The experts may entertain opposite conceptions, which would be difficult to reconcile, and be eager to over-refine the vocabulary.

The needs of future users must therefore be carefully assessed in order to perfect a well-proportioned tool to match them.

Determining the scope of the common portion will be a particularly delicate problem. The thesaurus must include all the reference-thesaurus concepts that can be of use in a specialised documentary system.

One could eventually endeavour to make the reference thesaurus so thin that it could be completely incorporated into each sub-thesaurus developed in parallel but we feel that this is unreasonable considering the thesaurus' present state.

---

[77] See note (14) in the second chapter.

Example

A retrieval aimed at PROGRAMMING LANGUAGE* ↓ will be "translated"

PROGRAMMING LANGUAGE "or" AUTHORING LANGUAGE "or" LOGO.

Various means are needed for compiling a sub-thesaurus : technical means (the thesaurus management software should be flexible enough to manage the unconventional elements, relationships and rules that we have described), human resources (to establish the content, perfect the data processing device, and manage the process) and funds (to produce and disseminate the documents).

Means must likewise be drawn upon to manage the feedback to the reference thesaurus[78] by coordinating the updating schedule for each of the two instruments.

Management of the sub-thesauri and of their reference thesaurus must therefore be **joint management**, both with regard to the tools (same methodology, same data processing system) and to the bodies in charge of the administration of these indexing languages (one administrator or group of administrators only).

Given this, how is it possible to make the most of the developments that have been achieved so far ?

We believe that the first stage of the work should involve the acknowledgement and assessment of the needs on the one hand, and the fine-tuning of the method and development of a data processing tool on the other hand.

It is indeed essential to go back to the potential users and question them in order to gather all the elements for a "workplan" that will guide those responsible for the compilation of sub-thesauri when they are faced with the above difficulties : determining the degree of specificity of the thematic coverage of the sub-thesaurus and limiting the scope of the portion common to the reference thesaurus.

Which specialised documentary systems would use the sub-thesaurus ? What are the -quantitative and qualitative- features of the documentary funds which they manage ? What means do they use to represent the content (classifications, indexing languages) ? Which ones do they recommend ? Which resources can they devote to the development of a tool which is meant for them (supplying of documentary sources, contribution of experts, funding) ? What need do they have to communicate with one another and with non-specialised documentary systems?

The answers to these questions will enable those responsible for the management of the Metathesaurus to decide whether it is worth creating sub-thesauri; to this end, the need expressed should be meticulously evaluated in terms of the costs involved , which must also take account of everything that is already available.

For example, the existing **"Vocabulaire destiné à l'indexation de documents en matière d'enseignement de langues vivantes en lien avec le thesaurus EUDISED"** [79] (trilingual vocabulary in English-French-German) could be relied on to organise the first stages in the process of establishing the content of a sub-thesaurus in the following way :

- appointment of a new group of experts with at least one representative per language;

- ratification of the wording of the terms (i.e. future sub-thesaurus descriptors) which have already been compiled and of the linguistic equivalences established so far : this removes the need for collection of the basic material, the firstselection and the establishment of language equivalents (provided that the work done beforehand is not fundamentally called into question and the idea of a sub-thesaurus in the nine languages is abandoned);

- ratification of the proposed groupings (approximately 900 compiled terms are divided between the existing microthesauri and the 25 or so proposed additional subdivisions).

---

[78] As a reminder, that feedback is concerned with addition, deletion and modification of concepts, structure modifications, marking the concepts which are anchoring points and introduction of non-preferred terms.

"Modification of concepts" means changes concerning descriptors, scope notes and non-descriptors.

[79] Council of Europe; prepared by Peter KNOPF; DECS/Doc (87) 24; Strasbourg; 1987.

The following stages are the same for the example above or for whichever sub-thesaurus it is agreed should be created :

- examining the semantic structuring of the concepts to be introduced into the sub-thesaurus (hierarchical and associative relationships) and their attachment to existing concepts;

- enriching these sub-thesaurus concepts by the creation of scope notes and introduction of non-descriptors;

- determining the grouping into nanothesauri;

- establishing proposals for the modification of the reference thesaurus;

- choosing the reference-thesaurus concepts which will appear in the sub-thesaurus.

Simultaneously, this method for designing and producing sub-thesauri will have to be ratified and a suitable data processing tool perfected.

We feel that the most natural way to proceed with ratifying that method would be to conduct pilot-projects within those specialised documentary systems which were involved in the compilation of the first sub-thesaurus.

As regards the data processing system, we have previously stated that it must be the same as that which manages the reference thesaurus. Its main specific functions must be : management of non-structured data which can be organised according to the complex structures described and the ability to offer sophisticated processing and publishing (control of the rules, publishing in nine languages, etc.).

3.4. SPECIALISED VOCABULARIES

3.4.1. Introduction

For some time now, those concerned with the compilation of a sub-thesaurus have had some inkling of the highly intricate and time-consuming nature of the work.

Determining the degree of specificity of the thematic coverage of the thesaurus is a problem which was encountered when creating the "NIT Extension " , which is a sectoral extension to the EUDISED Thesaurus to be incorporated into the next edition of the "European Education Thesaurus".

The experts in the field -the very people who had emphasised the need for supplementing the vocabulary so that it could be used in their specialised documentary systems- have not been allowed by those in charge of the thesaurus management (who are generalists aware of the needs of non-specialised documentary systems) to introduce all the concepts they wanted. An idea was then floated to draw up, independently of the thesaurus proper, "lists of specialised terms" whose status and possible use were not defined.

These two elements lead us to propose a form of specialised indexing language which is more rudimentary than a sub-thesaurus, but which can fulfil its main functions as regards the thematic cover and possibilities for communication.

3.4.2.Description and implementation

A **specialised vocabulary (SV)** is a set of lists (either mono- or multilingual) of specific terms. Each of these lists is attached to a concept in the thesaurus (which we shall also call reference thesaurus in this section).

Only one type of new element is to be created. We shall call it **specialised term (TS)** to indicate that, in terms of status, it is somewhere between a sub-thesaurus descriptor and a reference-thesaurus non-descriptor (non-preferred term).

Similarly, the **relationship** which links each specialised term to a reference thesaurus concept is somewhere between the hierarchical relationship and the semantic equivalence relationship. We therefore suggest the following symbol (see above) :

<DESCRIPTOR> UFS <SPECIALISED TERM>.

The only **rule** is that each specialised term must be linked by this relationship to one (and only one) concept in the thesaurus.

A specialised vocabulary does afford more detailed coverage of a restricted field and, at the same time, ensures the requisite possibilities for communication. The transition between two levels of indexing is performed by means of the new relationship in both directions (re-indexing or enlargement of a search statement to include the specific terms).

To decide whether a list of specialised terms should be **multilingual** or not, a choice will have to be made, in each case, between :

- the rapid creation of monolingual lists for restricted (or experimental) use;

- the establishment of language equivalence between terms from several languages, with the advantage of possible subsequent exploitation of the outcome in the context of the development of sub-thesauri.

The compilation and selection of the vocabulary, the establishment of language equivalences and the grouping into lists attached to reference-thesaurus concepts are indeed the first stages in the construction of a sub-thesaurus. The compilation and use of specialised vocabularies may in this case be used to check whether more important developments are useful.

As for the **methods of presentation** of the specialised vocabularies, we advise publishing separate documents rather than incorporting them into the reference thesaurus.

It would be possible to integrate the specialised vocabularies into the thesaurus (they could appear as non-preferred terms in the structured and rotated presentations, with a symbol which differentiates them from non-descriptors) but this would have major drawbacks :

- imbalance between the thesaurus language versions if specialised vocabularies are developed in monolingual versions or in a restricted number of languages;

- permanent feedback from all the work to the specialised vocabularies (to give an extreme example, each rehash of a specialised vocabulary would call for a new edition of the thesaurus) whereas their experimental character would not allow quick stabilisation.

Publishing separate documents, on the other hand, allows feedback to be limited in the editions of the reference thesaurus. The descriptors corresponding to reference-thesaurus concepts to which lists of specialised terms are attached can simply be "marked " by a distinctive feature, either in all language versions or not.

Specialised vocabularies then come in the form of relatively thin documents which can be republished quickly with each language version comprising :

- a "structured" presentation (remenbering, however, that there is only one hierarchical level) and a rotated pres entation of the specialised terms alone, with the indication of the descriptor corresponding to the concept to which they are attached (plus the possible language equivalences between specialised terms in the structured presentation);

- a presentation in alphabetical lists of the specialised terms under the corresponding reference-thesaurus descriptors, also alphabetically listed.

As far as implementation is concerned, the means to be drawn on are, relatively speaking, the same as for sub-thesauri, except for the necessary data processing tool which can be independent of the thesaurus management and perfected more rapidly.

CHAPTER FOUR : CO-THESAURI OR ASSOCIATED THESAURI

4.1. INTRODUCTION

Thanks to the sub-thesauri grafted onto the European Education Thesaurus, the Metathesaurus will be characterised by great flexibility, a feature which will enable it to keep up with developments in education and to meet the resulting, steadily growing range of needs for information exchanges and, hence, for documentary tools.

But its own **adaptation capacity** will not always suffice. In order to benefit from the specialised tools developed by different bodies, it should be accompanied by an **overture capacity**, so as to accommodate other indexing languages, i.e. existing thesauri.

Among the latter, two groups will be of special interest to those in charge of developing the Metathesaurus. They correspond to the following two cases :

a) an indexing language may be intended to cover the same field as the European Education Thesaurus. Such a language meets the same needs and should not, in principle, be used together with it on the grounds that (except for particular cases) a documentary system does not simultaneously use two methods to represent the content of the document which it itemises;

b) an indexing language may cover neighbouring fields, that is to say that its thematic coverage and that of the European Education Thesaurus may partially overlap. Such a language meets partially similar needs, and may be used

- together with the European Education Thesaurus in documentary systems specialised in a number of fields which are only properly covered if the two instruments are used together;

- alone in documentary systems whose speciality, very close to the field of education as it appears in the European Education Thesaurus, entails frequent relationships and information exchanges with documentary systems that use the latter (exchanges of bibliographical records, search statements addressed to both systems, ...).

a) cross-links

In the first case, the overture may be limited to establishing cross-links between thesauri with a view to helping regular users of the European Education Thesaurus when they need to use another thesaurus (for example, when they consult a data bank).

Each descriptor (concept) in the source thesaurus (i.e. the European Education Thesaurus) must be linked to the descriptor (concept) that is semantically the closest to it in the target thesaurus (e.g. the "UNESCO Education Thesaurus : IBE").

From a technical point of view a new type of relationship would be introduced -which could be called the (inter-thesaurus) "semantic proximity " relationship - between concepts in the source thesaurus and the corresponding concepts (or simply descriptors) in the target thesaurus, so as to provide the user with the elements necessary for re-indexing[80] a search statement composed with descriptors from the European Education Thesaurus, that is, express it in the indexing language used in the data bank which he is consulting.

---

[80] See above, 3.2.1. b), page 36.

This implies that an additional type of element is stored - namely, the corresponding descriptors in the target thesaurus (thesauri), which could appear in the structured presentation of the source thesaurus with an appropriate symbol [81] preceding them, or in "correspondence tables", published separately.

We shall not dwell on the various ways of locating the descriptors (concepts) which can be linked as described above[82], nor on the complexity of implementation, which increases if cross-links are to be established between multilingual thesauri[83], and demands suitable means (for instance for data processing).

---

[81] If A is a descriptor of the European Education Thesaurus and B the "semantically closest" descriptor in thesaurus T2, we shall have :

A 'T2 :'B

that will be read as

'to express' A 'in indexing language T2, use' B.

Example

TRADITIONAL EDUCATION **'IBE :'** CONVENTIONAL INSTRUCTION

[82] See RADA, R; *Connecting and evaluating thesauri : issues and cases*, Int. Classif. 14 (1987) N°2, p.63-69.

Examples

(built on the basis of EUDISED,1984 and IBE,1979)

1- Homographic or quasi-homographic descriptors :

PARENT RESPONSIBILITY **IBE:** PARENT RESPONSIBILITY

AUDIOVISUAL PROGRAMME **IBE:** AUDIOVISUAL PROGRAMMES

ADMINISTRATION OF EDUCATION **IBE:** EDUCATIONAL ADMINISTRATION

2- Descriptor of a thesaurus which is a non-descriptor in the other one

TRADITIONAL EDUCATION
UF CONVENTIONAL INSTRUCTION
IBE: CONVENTIONAL INSTRUCTION

3- Parallel hierarchical relationships of homographic or quasi-homographic descriptors : are the two specific descriptors of descriptor D1 in thesaurus T1 not likely to be "semantically close" to the two descriptors which in thesaurus T2 are specific descriptors of D2, homograph of D1 ?

EUDISED : LANGUAGES
NT1 CLASSICS
NT1 MODERN LANGUAGES

IBE : LANGUAGES
NT1 CLASSICAL LANGUAGES
NT1 MODERN LANGUAGES

4- It should be noted that particular cases are bound to be encountered :

* cross-link from one source descriptor to two descriptors of the target thesaurus:

SPORTS EQUIPMENT
UF SPORT FACILITIES
IBE: ATHLETIC EQUIPMENT
IBE: SPORTS FACILITIES

* no cross-link, i.e. no good "equivalent", is found in the target thesaurus; automatable procedures must then be replaced by expert appraisal which will permit a cross-link towards a more generic concept.

[83] It will have to be decided whether cross-links should be established between either corresponding language versions only or from all the versions of the source thesaurus towards only one version of the target-thesaurus, etc. .../...

Such an undertaking must obviously be justified by real and considerable needs (many users must frequently be faced with the need to go from the source indexing language to the target language).

Moreover, it lies far outside the scope of the context with which we are concerned. Easy access to items of information scattered among a wide variety of data banks is of interest to the server centres or international organisations that deal with these kinds of problems; they are the very ones that can gauge the large-scale needs and promote projects for sets made up of several indexing languages[84].

b) compatibility

The thesauri whose thematic coverages overlap with that of the European Education Thesaurus represent a different case, as they are **complementary** rather than "in competition".

If such complementarity is put into practice, i.e. if the thesauri are sometimes used either jointly or in isolation in documentary systems which often cooperate, their respective managers could set themselves the goal of making them **compatible**.

According to I. DAHLBERG [85], the compatibility of a thesaurus can be defined as a quality which permits :

1- the use of its own descriptors together with those of another thesaurus

or

2- the substitution of its own descriptors by those of another thesaurus.

---

[83] (cont.)
But multilingualism in the target thesaurus will help locate the semantically close concepts.

Example 3 in note (82) if non-descriptors and language equivalents are taken into account is:

EUDISED : LANGUAGES

NT1 E/ CLASSICS
UF **CLASSICAL LANGUAGES**
F/ LANGUES ANCIENNES
UF **LANGUES CLASSIQUES**
S/ **LENGUAS CLASICAS**

NT1 E/ **MODERN LANGUAGES**
F/ LANGUES VIVANTES
S/ IDIOMAS MODERNOS

IBE: LANGUAGES

NT1 E/ **CLASSICAL LANGUAGES**
F/ **LANGUES CLASSIQUES**
S/ **LENGUAS CLASICAS**

NT1 E/ **MODERN LANGUAGES**
F/ LANGUES MODERNES
S/ LENGUAS MODERNAS

[84] Like, for instance, the UNESCO project in the field of social sciences : see WHITELOCK, P.J.; A *descriptor bank of social sciences terms*, INT. Classif. 9 (1982) N° 3, p.145-151.

[85] In DAHLBERG, I.; *Conceptual compatibility of ordering systems*, Int. Classif. 10 (1983) n°1, p.5-8.

The proposed definition there concerns compatibility between "documentary languages" (read "ordering system" as classification language or indexing language): *"As an ordering system I regard any instrument for the organization, description (indexing) and retrieval of knowledge which consists of verbal or notational expressions for concepts and their relationships and which displays these elements in an ordered way, as e.g. a classification system, a thesaurus, a subject headings list or a similar device."*

"(...) compatibility of an ordering system is the quality which permits the elements of one such system to be used together with or interchangeably with the elements of another ordering system."

The second case (2-) corresponds to the "re-indexing" which will be necessary when documentary systems (using different indexing languages) cooperate in indexing (exchange of bibliographical records) or in information retrieval (search statements addressed to several documentary systems).

More accurately, it means in this case that :

- it must be possible to perform the substitution at descriptor level (i.e. re-indexing) while being sure that it is based on genuine equivalence between the concepts that those descriptors express;

- such a possibility is offered in all cases.

We shall refer to this situation as **"weak" compatibility**. It can be compared to the result obtained after cross-links have been established, but only if for all the source-thesaurus concepts the need to find at least one corresponding concept in the target thesaurus was respected (even if that concept is more generic in the fields where the source thesaurus offers more detail).

The first case (1-) corresponds to joint use of two thesauri in the same documentary system, which is only possible in a situation to which we shall refer as **"strong" compatibility,** for it includes one additional significant constraint : the equivalence of the *common concepts* must cause them to be expressed by identical descriptors.

As regards those concepts, re-indexing no longer applies; the hierarchical relationships between them may be either identical in both thesauri or not.

Apart from the common concepts, the choice of descriptors and the structure remain free, but a possibility of reindixing must be ensured (for example, by means of "cross-links" which will most of the time be established between concepts corresponding to very different levels of specificity).

The logic of more and more elaborate integration underlying the notion of compatibility may be developed even further, leading us to propose the notion of "association".

c) association

In the framework of the Metathesaurus, the thesauri "associated" to the European Education Thesaurus (which we shall call "co-thesauri" of the "reference thesaurus") will be "compatible" with it, although the meaning of the word is here more precise than that it was in the dual definition above. The rules governing association will be more limiting than those applying to compatibility :

- in the common fields, joint use of both thesauri is based on the identical nature of the concepts, of the descriptors which express the concepts, and of the semantic structuring of the concepts;

- the need for re-indexing is limited to the fields which one of the thesauri covers in more detail than the other, but it must be made possible by a strong link : an attachment of the sub-thesaurus-type.

In other words :

- the co-thesaurus is identical to the reference thesaurus in areas which both cover to the same level of specificity;

- the co-thesaurus has the same features as those of a sub-thesaurus in areas that it covers in more detail than the reference thesaurus. Conversely, the reference thesaurus will be like a sub-thesaurus of the thesaurus which is associated to it in areas that the associated thesaurus covers in more detail than the reference thesaurus;

- there should be no area totally specific to either associated thesaurus. To make re-indexing possible in all cases, the top of all the hierarchical chains must belong to both thesauri [86].

A graphic representation of this conception is provided in Figure 9.

To tackle the general problem of the **close** (or more specialised) **thematic coverage** provided by indexing languages other than the European Education Thesaurus with which the **possibility of communication** is sought, we have at our disposal a series of "linking" methods that gradually become narrower :

- **cross-links** are the first means; they are established between indexing languages that remain largely independent; re-indexing is not necessarily made possible;

- **compatibility** demands that re-indexing be made possible in all cases; its **"weak"** variant allows equivalent concepts to be expressed by different descriptors, whereas its **"strong"** variant does not;

- **association** imposes the need for a genuinely identical character (in the semantic structuring as well) in the common fields and a method of attachment based on strong constraints in the specialised fields specific to each indexing language.

The definition which we have put forward for co-thesauri corresponds to a very narrow linking method and relies on the features of sub-thesauri. It is therefore obvious that association will only be possible between indexing languages which are, largely **managed on a joint basis.**

As such, this definition does not require any developments to complement those set out in the previous chapter on sub-thesauri; instead, we propose to concentrate on the **method for carrying out** approximation, as one of the essential features of association is the fact that it relates to existing thesauri.

There are few pairs of existing thesauri which exactly meet all the conditions for eventual association. The following four points summarise these conditions :

a) the thematic coverages overlap;

b) the thesauri are used in documentary systems whose "vocation" is to have frequent relationships;

c) their structures are analogous (same types of elements, relationships, and (preferably) multilingualism)[87];

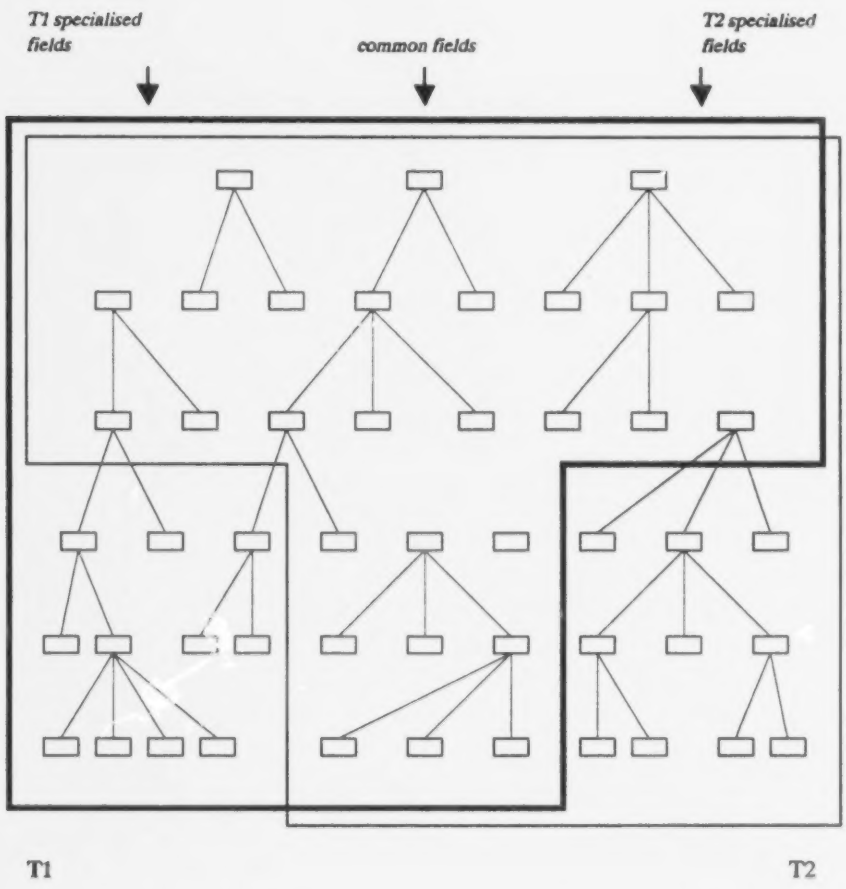
d) those in charge of the thesauri management are willing to approximate the thesauri to be associated, and "co-management" can be organised.

---

[86] See above, 3.2.3. a), page 38 (rule 3).

[87] But, at this stage, their semantic structuring is not analogous.

Figure 9. Co-thesauri



The CEDEFOP "Thesaurus of vocational training" (we shall refer to it as "the CEDEFOP Thesaurus" [88]) and the European Education Thesaurus meet these requirements. There is a will to approximate the thesauri, the need for such approximation has been acknowledged, and after declarations of favourable intent for what was termed the "search for compatibility" or "concern with greater coherence", the first concrete endeavours were made when the EUDISED Thesaurus was being updated and to become the European Education Thesaurus.

This example will give impetus to our investigation into the approximation process, the different stages of which correspond to some extent to the ever narrower "linking" methods introduced above.

4.2. APPROXIMATION PROCESS

4.2.1. Conditions

The will to approximate expressed by those in charge of managing the European Education Thesaurus and the CEDEFOP Thesaurus originates in and is mainly justified by the obvious complementary nature of education and vocational training, which are more and more closely united in practice as well as in national and European policies.

We shall demonstrate that in addition to this fundamental argument, all the conditions necessary for the envisaged undertaking have been met.

a) There is a genuine **overlap of the thematic coverages** of the two indexing languages. This was taken into account as soon as the development of the CEDEFOP Thesaurus started (the Introduction to the first edition (1986) mentioned the EUDISED Thesaurus among the indexing languages referred to), and it has justified and fostered the ongoing dialogue.

b) Cooperation bonds and analogous aims, structures and ways of functioning unite the two information networks on education and vocational training in the European Community, namely, the EURYDICE and CEDEFOP networks.

The corresponding **documentary systems** in the Member States of the European Community and those that exist at community level -both privileged users of these thesauri- do indeed have a "vocation" to cooperate.

c) The **analogy of structure** between the two thesauri (types of elements and relationships) is almost complete: the CEDEFOP Thesaurus only differs from the EUDISED Thesaurus as we have described it in that it has no microthesauri (and hence no graphic presentation).

But this difference reflects another : its vocabulary is organised into few hierarchies which are very heavy. Whereas the height of the 243 hierarchical chains in the EUDISED Thesaurus seldom exceeds three levels, the 19 hierarchical chains in the CEDEFOP Thesaurus often comprise seven levels[89].

---

[88] See above, 1.3.3., page 5.

[89] The problem of grasping the vocabulary in its entirety has been resolved by the creation of a type of presentation which allows the top of all the hierarchical chains to be visualised : the alphabetical list of the 19 top terms supplemented with their 127 directly specific terms (see Annex 4) allows the retrieval of a descriptor to be directed in a way similar to that used for consulting the list of the 42 microthesauri and locating the relevant top terms among the 243 existing ones which is made easier by their prominent position in the graphic presentation of the EUDISED Thesaurus.

Moreover, polyhierarchies are frequent i n the CEDEFOP Thesaurus and they are not limited to the most specific concepts [90]. This makes it even more difficult to reconcile the respective semantic structures of these two thesauri.

As for multilingualism, the CEDEFOP Thesaurus has seven language versions [91] whereas there are nine in the EUDISED Thesaurus.

d) In addition to the **will to approximate** already mentioned, there are also several favourable elements which could provide possibilities for concrete organisation of the **joint management.**

Let us first remind the reader that both thesauri have the same father, the late Mr. J. VIET, author of the first edition of the EUDISED Thesaurus and co-author of the second, who prepared the first edition of the CEDEFOP Thesaurus.

The association of those in charge of the CEDEFOP Thesaurus management at all stages of the content updating of the future European Education Thesaurus since 1986 and the experience of the first joint work meetings have demonstrated that even deeper cooperation was possible and that it is of outstanding interest to each of the parties involved.

Lastly, the fact that the same management software (ASTUTE [92]) has so far been used would enable the two thesauri to migrate in parallel towards a new common management system, as the need for the same data processing device is another consequence of the type of association defined above [93].

---

[90] **Example**

PLANNING OF TRAINING

BT1 EDUCATIONAL PLANNING
BT2 EDUCATIONAL ADMINISTRATION
BT3 EDUCATION
BT4 EDUCATION AND TRAINING
BT1 TRAINING ADMINISTRATION
BT2 TRAINING
BT3 EDUCATION AND TRAINING
BT1 TRAINING POLICY
BT2 TRAINING
BT3 EDUCATION AND TRAINING

NT1 COURSE DESIGN
NT1 TRAINING EVALUATION
NT1 TRAINING MARKET
NT2 TRAINING NEEDS
NT2 TRAINING SUPPLY
NT2 TRAINING-EMPLOYMENT RELATIONSHIP
NT1 TRAINING STATISTICS
NT2 NUMBER OF TRAINEES
NT2 TRAINING NEEDS ANALYSIS

It should be noted that such practice causes considerable data redundancy in the structured presentation which, owing to the bulkiness of each hierarchy, is incompatible with a graphic presentation but which could be regarded as a way of increasing opportunities to locate the relevant parts of the hierarchical chains.

[91] The CEDEFOP Thesaurus was initially compiled in three languages : English, French and German. The Italian version was added to the first edition whereas the Greek, Spanish and Portuguese versions were added to the second edition, which only differs from the first one in this respect.

[92] See note (9) in the first chapter.

[93] For instance, the MICT software (see note (10) in the first chapter), which has been adopted for the next edition of the European Education Thesaurus.

We shall present the approximation process in two phases.

The first corresponds to an approach focussing on the concepts (descriptors, non-descriptors and scope notes) and aims to meet the following dual requirement : the same concepts are expressed by the same descriptors and identical descriptors (homographs) express the same concepts.

The second phase corresponds to enlarging the process to include the hierarchical chains and must result in determining the composition of the common hierarchies and the respective limits of the two thesauri in the areas where one appears as a sub-thesaurus of the other.

4.2.2. Localised approach

The aim of the first phase is to ascertain and, if need be, establish exact correspondence between the forms for expressing of the common concepts (the descriptors) in each of the indexing languages to be associated.

More accurately, what must be established is their *unique* form of expression in the portion that will be common to the two indexing languages when they are associated.

We split this phase into two stages corresponding to the following two cases :

- concepts that can be approximated on the basis of similarity between descriptors and/or non-descriptors;

- concepts that can be approximated because of similarities between their hierarchical positions with reference to those of common concepts identified during the first stage.

The automatable procedures alluded to are reminiscent of the methods for locating "semantically close" concepts in two thesauri between which cross-links are to be established[94].

a) first stage

The corresponding procedure consists of three steps, the first of which can be automated.

1- Identify the concepts

- expressed by descriptors which are, in one or several languages, homographs or quasi-homographs;

- expressed in one thesaurus by descriptors which are, in one or several languages, homographs or quasi-homographs of non-descriptors in the other thesaurus.

2-Ascertain the equivalence of the underlying concept. This equivalence is assessed in considering all the elements that go towards expressing the concept : not only the wording of descriptors but also all the non-descriptors and possible scope notes, in all the common languages, and the direct hierarchical environment.

3-If conceptual equivalence is established, choose the unique form of the expression : the wording of the descriptors in each common language.

Examine in each case whether it is worthwhile to conserve as non-descriptors the wordings which were not retained (former descriptors in either of the thesauri), to re-write the scope notes on the basis of the existing ones or to create them and keep all the non-descriptors.

---

[94] See notes (82) and (83).

If equivalence is not established, the concepts are different (but probably close [95]).

In this case, it must be ensured that no descriptor or non-descriptor whatsoever is common to the two concepts and, where necessary, scope notes must be adapted or created to remove possible ambiguities and mark the distinction better.

It is worth noting that at this stage we are not yet dealing with the final decision as to which concepts will be common to both associated thesauri and which will be specific to one or the other. This is the aim of the second phase (approach enlarged to include the hierarchical chains).

b) second stage

The second stage in the first phase (approach localised to concepts) consists of processing all the concepts that were not processed in the first stage: those that the above procedure has not approximated as possible equivalents. The aim is to locate the concepts which are nevertheless equivalents, despite their having no identical or similar descriptors or non-descriptors.

The corresponding procedure can be compared to that above; automatic processing facilities may be used in the first step by resorting to the hierarchical relationships to approximate the concepts which are, in each thesaurus, directly generic or specific to the common concepts identified during the previous stage.

1- Systematically **approximate** the concepts whose hierarchical position is the same compared to a common concept.

2- **Establish possible equivalence** not only between these concepts but also between all those that other explorations of the two vocabularies -which might be more intuitive than systematic- could suggest.

3- **In the case of equivalence,** proceed as before. Choose the unique form to express the common concept (the descriptors) and take decisions as to the wordings that were not retained, the non-descriptors and the scope notes.

4- **Reiterate** the procedure to extend the investigation to the generic and specific terms of each new common concept established.

The result obtained by that procedure corresponds to one of the two conditions for "strong" compatibility as previously defined[96]. The common concepts are expressed identically, a fact which will permit joint use of the two thesauri.

The second condition (possibility of re-indexing in all cases) can nevertheless only be fulfilled by extra work, which is very likely to be considerable, on establishing the necessary cross-links according to details on which we shall not dwell here, as association, which is our aim, provides a qualitatively superior solution to that problem.

---

[95] It should be noted that the automated procedure will approximate concepts which are even united by such a flimsy link as inexact homography between a pair of non-descriptors in one language.

The automatically emphasised link could be weighted so as to order the approximated couples of concepts on a scale ranging from the strongest to the weakest link according to the idea that the actual equivalence probability between the expressed concepts is at its highest for exact homography of the descriptors in all the common languages and diminishes as the number of differences increases (such balancing will only provide an indication and will never be a criterion for equivalence).

[96] See above, 4.1. b), page 51.

4.2.3. Enlarged approach

At the end of the first phase, three sets of concepts were distinguished - namely, the concepts that are common and those that are *a priori* specific to each of the thesauri to be associated. The final content of these three sets must be decided on.

We have also partially anticipated the second phase in the course of the second stage since similar parts of hierarchical chains (common concepts identically structured on two or more levels) were very likely given prominence and since the main differences in structure were located.

Reconciling the respective hierarchical structures of the two thesauri to be associated is a complex task whose progress cannot be formalised as clearly as that of the first phase.

Most of the considerations relating to the consequences of producing sub-thesauri could be repeated here[97]. The need for largely joint management has already been mentioned and considerable feedback (which is, as said before, also useful) must be taken into account. The features of two thesauri resulting from association may well be very different from those they had before the approximating process.

For instance, it is obvious that in the philosophy underlying the general structure, in a case where the thesauri to be associated are as disparate as the CEDEFOP Thesaurus and the European Education Thesaurus, one of the two ideas must be privileged.

In our example, if one wishes to maintain the option of a graphic presentation in the form of terminographs, the thesaurus of vocational training resulting from the association must also be organised into microthesauri and more numerous and less loaded hierarchical chains, and the polyhierarchies must be limited to the most specific concepts (such constraints are not major obstacles[98]).

Conversely, as regards the establishment of the hierarchical relationships considered "locally", organization of the work must be guided by a principle of equality :

- groups of specialised experts may be given the task of (re)building the hierarchical chains and establishing the respective limits of the two thesauri in the specialised fields covered by each indexing language;

- in the common fields, "co-management" is put into practice by joint working meetings.

Lastly, it should be stressed that the approximation process must be planned over a relatively long period :

- the work procedures and the corresponding organisation must be set up;

- the complexity and challenges of the undertaking, demand that much thought be given to decisions on content;

- each of the thesauri to be associated must remain operational. It is worth providing for periodic updatings if they occur frequently enough.

---

[97] See above, 3.3., page 44.

[98] The EUDISED Thesaurus was similarly transformed between the first and the second edition.

Sub-thesaurus and co-thesaurus are based on the same principle of attachment which, for data processing management especially, will only require one adjustment.

Sub-thesaurus and co-thesaurus come within the scope of the same project i.e. they endow with a "common language" all those involved in European cooperation in education in the broadest sense of the term, that is including specialised fields and neighbouring areas, such as vocational training.

In the event of an association between the European Education Thesaurus and the CEDEFOP Thesaurus, the Metathesaurus will be the set of indexing languages that centre on a dual core : each of the associated thesauri will be the reference thesaurus not only of sub-thesauri but also of "national vocabularies".

CHAPTER 5 : NATIONAL VOCABULARIES

5.1. INTRODUCTION

Attaching the various types of indexing languages introduced above to the same reference thesaurus, the European Education Thesaurus , aims to ensure, through that common language, a communication possibility between documentation centres whose variety is in keeping with one of the two characteristic main lines of the field to be covered i.e. the **interdisciplinarity** of the science of education.

Consequences of this feature are notably that its boundaries with the other disciplines are blurred (hence the relationship with thesauri covering neighbouring fields is so very important), and that many specialised disciplines are concerned (hence the importance of being able to cover them specifically with sub-thesauri).

But the other characteristic main line to which we have given prominence throughout[99] is at least as important from the angle of the European project. It is the **internationality** of the educational situation.

The "national" particularities or, more exactly, the variety of educational systems, have a corresponding variety of vocabularies that is both an obstacle to communication which must be removed for documentary purposes and a source of information on educational situations -their analogies, their differences- which can be exploited in work on the fringe area of actual documentary concerns.

A preliminary remark is necessary here. In the expression "national vocabularies", the term "national" is in fact incorrect since there are more European educational systems than countries (the United Kingdom, for instance, has several different educational systems).

Nevertheless, for greater convenience (and in deference to an expression which has come into current use), we shall call "national" vocabulary what is in reality a vocabulary "specific to an educational system".

It should also be noted that the vocabulary describing the actual European situation is made more complex by the relation between educational systems and language areas :

- a language area may encompass several educational systems (for instance the Belgian, French and Swiss educational systems in the French-speaking area);

- an educational system may apply to several language areas (for instance, the Swiss educational system in the German-, Italian- and French-speaking areas);

- in the other cases, the two sets correspond (for instance, the Spanish educational system corresponds to the Spanish language area in Europe).

Some bodies responsible for information on education at national level have started compiling and sometimes structuring terminology specific to their environment, which they need to represent accurately the content of the documents being itemised.

---

[99] See above, 1.3.1., page 3.

In the fringe area of work on the European Education Thesaurus, the two greatest achievements have so far come from France and Austria :

- a list of Austrian "national descriptors" was published by the Council of Europe [100];

- a list of "additions to the EUDISED Thesaurus" was prepared by the Centre National de la Documentation Pédagogique (Paris) [101].

Hereafter, the terms "national vocabulary" will refer to a form of structured indexing language covering the specific educational factors of a given educational system.

There are several reasons why intrest is being shown in national vocabularies in relation to the Metathesaurus.

a) **possibilities for communication** remain central.

National vocabularies are used as complements to the European Education Thesaurus and together with it, so as to reach a higher level of accuracy when indexing documents with a national character (they would, for instance enable direct retrieval of all the documents dealing with CIOs (Centres d'Information et d'Orientation) in a French documentary system).

Where information is shared with the users of the reference thesaurus (used either alone or with another national vocabulary), the possibilities for communication must be ensured by means of an attachment that allows re-indexing of that part of the indexing which was carried out with the national vocabulary into the "common language" (CIO will be "translated" into the descriptor SERVICE D'ORIENTATION (GUIDANCE SERVICE) when re-indexing takes place).

b) It is also important to **disseminate** these vocabularies.

These vocabularies whose main purpose was at first largely national deserve a wider audience, for the following three reasons :

- some documentary systems -notably at community level- itemise documents from different countries [102];

- national vocabularies will also be useful in the European information networks ;

- a tool whose use is not limited to strictly documentary matters will be available, since researchers in comparative education could also use it. We shall come to this later [103].

---

[100] Council of Europe; *EUDISED Thesaurus : Appendix 1 : Austrian national descriptors;* list prepared by the Austrian Ministry of Education, Culture and Sports; DECS/Doc (89) 7; Strasbourg; 1989.

[101] CNDP; *Liste des additifs au thésaurus EUDISED : Document de travail : Propositions pour l'adaptation du thésaurus EUDISED aux particularismes du système éducatif français;* not published; 1987.

[102] In the case of documentary systems itemising a large number of documents of miscellaneous national origin and devoted, for instance, to legal rules relating to educational structures in different countries, it will be noticed that the discriminating power of the European Education Thesaurus descriptors is not sufficient for practical usage : the combined use of several national vocabularies is then advised.

[103] See below, 5.2.5. b), page 76.

c) **Feedback** to the European Education Thesaurus is desirable.

The thesaurus can be improved during systematisation, structuring and attachment operations which are likely to throw up many new questions and suggestions, reflecting the diversity of educational situations and thus permitting the "European dimension" of the thesaurus to be strengthened.

Furthermore, the existing national vocabularies have proved to be a privileged source of proposals for new descriptors in updating the content of the European Education Thesaurus. It is the primary and most direct form of feedback.

For all these reasons, it seems wise to coordinate, from a methodological point of view, the initiatives from the various national parties who are developing or want to develop structured vocabularies covering their specific educational situations.

This can be achieved if we offer them a common work frame - a definition of the structure of the national vocabularies and of their method of attachment to the European Education Thesaurus (which remains the "reference thesaurus").

The constraining character of this "work frame" is necessary to reach the triple goal which we have set ourselves. The goals relating to the specialised thematic coverage and possibilities for communication -which are common to all developments proposed in the framework of the Metathesaurus- are accompanied, in this case, by the will to exploit the precious source of information from what can be regarded as "coordinated terminologies" as long as the procedural steps are identical.

But limitations should go together with incentives. The functional nature of better communication is the first such incentive. Nevertheless, the appropriate tools still need to be available.

This is why we think that the Metathesaurus managers must offer logistic aid (software and means to publish and disseminate the vocabularies) so as to further the development of national vocabularies.

5.2. DESCRIPTION

5.2.1. Aim

a) thematic coverage

National vocabularies must answer the needs of the documentary systems which itemise literature of a mainly national character.

Only a few terms matching these needs can be found in the European Education Thesaurus since the applied selection criterion first rejected the concepts that were not common to all the educational systems and was then altered to permit the introduction of concepts unknown in one country or one language area [104].

The debates caused by enforcing the amended rule have shown the difficulty and limits of such a practice. But these debates were also unquestionably fruitful; they fit in perfectly with plans to improve *mutual knowledge* of the different European educational situations.

This was an opportunity to realise that work apparently so remote from concrete action, namely the development of documentary tools, may help solve the very first problem encountered in any European undertaking : *mutual understanding*.

---

[104] See above, 2.2.1., page 12.

A national vocabulary will comprise (simple or compound) terms in use in a given educational system; they may exist in several languages (if the educational system has more than one official language).

We shall divide up these terms into three categories :

- terms expressing a concept specific to a particular **educational system**, which has no equivalent in **all** the other educational systems (example : the "Local Education Authority" in United Kingdom);

- terms expressing a concept common to a **language area**, which has no equivalent in **some** of the other languages (example : the notion of "animateur" is unknown in English-speaking countries);

- **proper nouns** : names of institutions and people (e.g. : the French Ecole Nationale d'Administration and the educationalist Ovide Decroly).

In the first two cases, a translation may still be possible, albeit by circumlocution. It may be *more or less* difficult to grasp the concepts concerned and express them in a given language.

We have stressed that they are concepts which are not common to **all** the educational systems or **all** the language areas. In fact a continuum extends from the common concepts to the particular concepts, which can be marked out as follows :

1- concepts common to all the educational systems and/or all the languages;

2- concepts unknown to only one educational system but which could be validly expressed in all the languages;

3- concepts difficult to express in one language (this exception corresponds to one or several educational systems, depending on the case);

4- concepts unknown in a minority of educational systems and/or language areas;

5- concepts unknown in a majority of educational systems and/or language areas;

6- concepts specific to only one language area (which corresponds to either one or several educational systems, depending on the case);

7- concepts specific to only one educational system (one or several language areas are concerned, depending on the case).

A final note is that the thematic coverage of the national vocabularies must be regarded as *limited to certain fields* rather than "transversal" compared to that of the European Education Thesaurus. Almost half of the terms appearing in the two examples mentioned (Austrian and French national vocabularies considered as a whole) are attached to the following three micro-thesauri [105] :

---

[105] The table below mentions the nine microthesauri to which most of the national terms are attached in the Austrian (VNA) and French (VNF) national vocabularies. These microthesauri are the ones to which more than 7% of the terms from either of the national vocabularies are attached (where no figure is provided, it means that there are less than 7% of terms are attached).

Columns 1 and 2 : serial numbers and titles of the microthesauri.
Column 3 : number of terms attached in the Austrian national vocabulary.
Column 4 : corresponding percentage (calculated on the basis of the total amount of terms in the Austrian vocabulary, i.e. 146).
Column 5 : number of terms attached in the French national vocabulary.
Column 6 : corresponding percentage (calculated on the basis of the total amount of terms in the French vocabulary, i.e. 244).
Column 7 : total number of attached terms for the two national vocabularies (VNA + VNF).
Column 8 : corresponding percentage (calculated on the basis of the total amount of terms in the two vocabularies, i.e. 390). .../...

(01) TEACHING AND TRAINING
(04) SYSTEM OF EDUCATION
(34) PROFESSION AND PERSONNEL

b) possibilities for communication

These are ensured by the method of attachment of the national vocabularies to their common reference thesaurus, i.e. the European Education Thesaurus.

In principle, the concepts common to all the systems and/or all the languages (case 1- in the above scale) are present in the reference thesaurus.

As previously stated, national vocabularies are stocks of "candidate descriptors". They are very good indicators of the actual needs insofar as they are managed independently and can thus be adjusted more rapidly to these needs.

The terms which can unmistakably be recognised as "national terms" (case 7-) will never become descriptors in the reference thesaurus.

Between these two extremes lie all the other terms. These correspond to concepts about which no *a priori* assessment can be made by those responsible at national level of the extent to which they are either specific to their environment or common.

For these terms, the procedure that organises investigation and debate on proposals for new descriptors with all those involved in the reference thesaurus management consists of locating them on the scale of 1 to 7.

* They are very likely to be accepted if they turn out to belong to case 1.

* They may be accepted after more tougher debate if they correspond to cases 2 or 3 (the "loan-word" technique may be resorted to for some language versions[106]). In cases 3 to 7, they may become non-descriptors in the reference thesaurus, which does not prevent them from also being "national terms" in the national vocabularies.

  Some of these terms express existing factors common to at least one language area (whatever the number of educational systems concerned, cases 2 to 6) and can, as such, be regarded as non-descriptors referring to reference-thesaurus descriptors .

  In principle, transnational cooperation between partners belonging to the same language area must allow different language versions of the thesaurus to be supplemented with such non-descriptors.

---

[105] (cont.)

	VNA		VNF		VNA+VNF	
	nb	%	nb	%	nb	%
(01) TEACHING AND TRAINING	-	-	61	25	65	16.6
(04) SYSTEM OF EDUCATION	43	29.4	18	7.3	61	15.6
(05) ADMINISTRATION OF EDUCATION	10	6.8	22	9	32	8.2
(06) EVALUATION	18	12.3	17	6.9	35	8.9
(09) CURRICULUM	18	12.3	-	-	24	6.1
(27) INTERRELATIONS	11	7.5	-	-	17	4.3
(34) PROFESSION AND PERSONNEL	18	12.3	40	16.3	58	14.8
(35) ADMINISTRATION	-	-	18	7.3	18	4.6
(36) PUBLIC ADMINISTRATION	-	-	22	9	24	6.1
Total amounts (all the microthesauri)	146	100.0	244	100.0	390	100.0

[106] See note (18) in the second chapter.

These partners may also introduce case 7 terms as non-descriptors with a view to multiplying the access points from natural language of documents and questions into such a controlled language as a documentary language.

We have just mentioned the possibility of intregrating some "national terms" into the reference thesaurus as either descriptors or non-descriptors.

Remember that the status of non-descriptor and that of "national term" are compatible but the status of descriptor and that of "national term" are not. A "national term" which becomes a reference-thesaurus descriptor disappears from the national vocabulary from which it came.

Indeed, unlike the sub-thesauri, there will be no concepts common to both systems : the reference thesaurus and the national vocabularies will be presented as two disjointed sets of concepts. We shall now tackle the issue of the linking method to be established between the two systems which must allow for the possibility of communication.

As the sub-thesauri, the hierarchical relationship will form the channel towards the "common language". Such a channel will be crossed by a "re-indexing" operation [107].

The concepts expressed by "national terms" will be differentiated according to the way in which they may or may not be linked to the reference-thesaurus concepts by a hierarchical relationship.

- ***hierarchy***

  The concepts which may be considered as specific to reference-thesaurus concepts are linked to them by a direct or indirect hierarchical relationship (because they are structured into hierarchical chains, allowed for by the national vocabulary).

  The corresponding rules will be dealt with later[108].

- ***polyhierarchy or compulsory multi-equivalence***

  Concepts which can only be validly expressed by a combination of reference-thesaurus concepts, for which a mere hierarchical relationship would be far too narrow, could be linked by a polyhierarchical relationship.

  Such a solution is not completely satisfactory. We doubt whether polyhierarchy is the clearest way of representing the need for a combination of two or several concepts.

  Some thesauri allow a type of relationship that is closer to that idea called **compulsory multi-equivalence.**

---

[107] See above, 3.2.1. b), page 36.

[108] See below, 5.2.3., page 70.

[109] **Examples** (from the CEDEFOP Thesaurus)

* Italy may be viewed either as a Mediterranean country or as a Member State of the European Community :

ITALY
BT1 EEC COUNTRIES
BT1 MEDITERRANEAN COUNTRIES

* Delinquency may be viewed as an individual or social problem:

DELIQUENCY
BT1 ANTISOCIAL BEHAVIOUR
BT2 SOCIAL BEHAVIOUR
BT3 BEHAVIOUR
BT1 SOCIAL PROBLEMS
BT2 SOCIETY

Figure 10. Compulsory multi-equivalence relationship

*** Its interest**

The compulsory multi-equivalence relationship (sometimes known as "compound semantic equivalence") strengthens the links established between the documents' natural language and the thesaurus' controlled language : its use increases the indexing coherence, i.e. the likelihood that one subject is expressed by means of the same descriptors by different indexers.

In the case of compound terms whose components are already descriptors, multi-equivalence has not much to contribute as most indexers will choose the descriptors

WORKING CONDITIONS and TEACHER to index a document on "teachers' working conditions".

It should however be noted that it is not always possible to compare situations in a multilingual thesaurus' different language versions.

On the other hand, the term SOCIO-EDUCATIONAL EQUIPMENT is not necessarily understood by everybody as "the buildings and aids for cultural activities". Even if it were, it would probably be more adequate to guide the indexer towards the following three descriptors : EQUIPMENT, BUILDING and SOCIO-CULTURAL ACTIVITIES.

*** Symbols**

- If A is the non-descriptor and if B and C are the corresponding descriptors, the symbols are :

A 'USE' B 'AND' C or A 'USE' B '+' C

which can be read as

'to express' A 'use' B 'in combination with' C

Example TEACHERS' WORKING CONDITIONS
USE TEACHER AND WORKING CONDITIONS

- To make the relationship appear in the structured presentation under the entries of each of the descriptors concerned, either of the following symbols may be used :

B
+C UF A

C
+B UF A

B
UFC A (+C)

C
UFC A (+B)

("Used For Combined")

Example

BUILDING
UFC SOCIO-EDUCATIONAL EQUIPMENT
(+EQUIPMENT)
(+SOCIO-CULTURAL ACTIVITIES)
...
EQUIPMENT
UFC SOCIO-EDUCATIONAL EQUIPMENT
(+BUILDING)
(+SOCIO-CULTURAL ACTIVITIES)
...
SOCIO-CULTURAL ACTIVITIES
UFC SOCIO-EDUCATIONAL EQUIPMENT
(+BUILDING)
(+EQUIPMENT)

This relationship, a variant of the semantic equivalence relationship, links one non-descriptor to two (or several) descriptors that must be used together (see Figure 10).

Polyhierarchy is confined to multiplying the access channels to a specific concept which may be considered from different angles[109] whereas multi-equivalence emphasises that a combination of concepts is essential to grasp and express a given concept correctly while respecting its wealth and uniqueness.

In order to use this relationship as a method of attachment, it must be admitted that, in the national vocabulary, the term to be linked in that manner is definitely a preferred term, which can be validly used for indexing and which can have hierarchical relationships with specific terms.

This does not predetermine whether it will become a reference-thesaurus non-descriptor or not. Such a possibility is subject to the condition that the compulsory multi-equivalence relationship be introduced in the reference thesaurus (in the common acceptation of the phrase this time, i.e. as a relationship between non-descriptors and descriptors). Besides, this practice corresponds to the wish expressed by some parties involved in the ongoing management of the European Education Thesaurus [110].

The use of the compulsory multi-equivalence relationship as a way of attaching concepts specific to one educational system to the "common-language" vocabulary will be encouraged as it is the best-suited for expressing the various underlying realities in as accurate and non-reductive a way as possible.

Its use will allow :

- greater accuracy in the indexings and efficiency of the search statements which must be re-indexed (on certain conditions with which we shall deal later [111]);

- an awareness of the analogies between concepts specific to different educational systems which are however represented by a combination of the same "common-language"concepts [112].

- ***no relationship***

The concepts that cannot be attached by the means listed above (i.e."isolated" concepts) must be allowed to appear, if only provisionally, in the national vocabularies, for their presence may form interesting items of information

- on how the situation and conceptions evolve in the different educational systems;

- on the possible gaps in the reference thesaurus[113].

---

[110] Some users would appreciate better guidance when they encounter subjects that require the combination of several descriptors for indexing.

On the other hand, we have met users who view indexing as one of the intellectual acts of the documentalist's profession and consider that the documentalist should remain free to make the relevant choices.

We think that post-coordination recommendations could certainly enhance the coherence of indexings of a miscellaneous origin like those met in locally fed data banks.

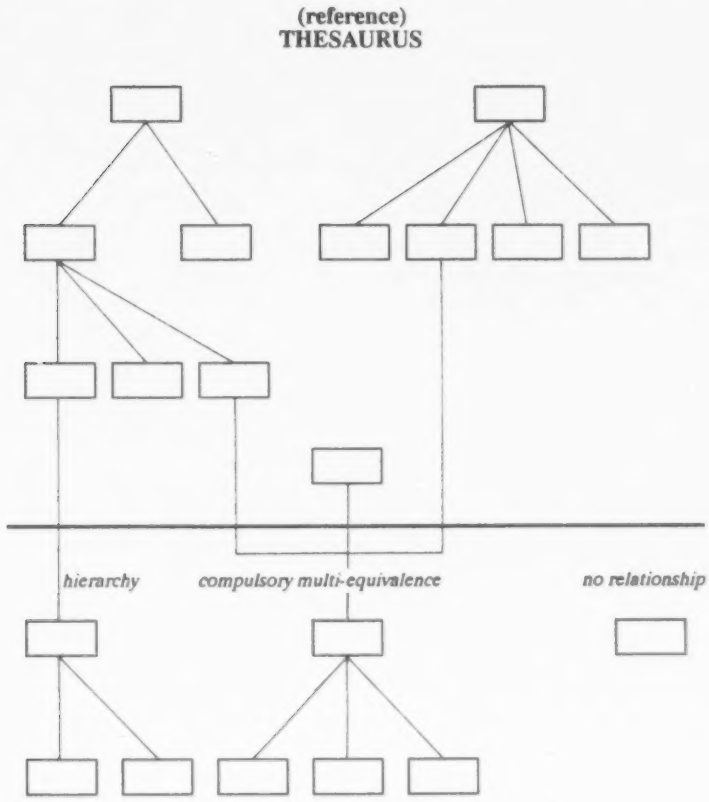
But we also think that, even in the case of non-descriptors, the establishment of each multi-equivalence relationship to be introduced should be the result of multilateral dialogue leading to a common decision, which could result in introducing a (*non-compulsory*) linguistic equivalence relationship between non-descriptors.

[111] See below, 5.2.5. a), page 74.

[112] See below, 5.2.3. e), page 72 and 5.2.5. b), page 76.

[113] The fact that educationalists' names are present in their national vocabulary has led the French to propose the introduction of the descriptor EDUCATIONALIST in the European Education Thesaurus to fill a gap which was revealed because it was impossible to find a suitable generic term.

Figure 11. The national vocabularies' methods of attachment



These various cases are represented in the form of graphs in Figure 11.

5.2.2. Elements

The **"national vocabulary"** (NV) system only comprises two basic elements : the **national terms** (TN) (whose function is analogous to that of descriptors) and the **non-preferred national terms** (NTN) (analogous to non-descriptors).

We suggest imposing a **definition** to be appended to each national term instead of waivering over the possibility of adding a scope note.

The aim of this rule (which was respected by the compilers of the Austrian national vocabulary) is to encourage a systematisation process which could be exploited in other ways and to contribute, through the dissemination of national vocabularies, to better mutual knowledge of the various educational situations among partners from different educational systems.

When they work with the national vocabulary that corresponds to their own educational system, the users will seldom need details on the acceptation or usage of a particular national term. Hence they will seldom need a scope note; but the definition will be an invaluable source of information for users from other countries.

We do not make the definition a distinct element, so emphasising that it is a *compulsory attribute* of each national term.

5.2.3.Relationships, rules and subsystems

To distinguish the relationships used in the national vocabularies, we shall distribute them into five non-disjointed sets :

a) relationships that organise the subsystems "national concepts" : semantic equivalence and linguistic equivalence;

b) relationships that organise the structure specific to the national vocabulary : hierarchy, association and inclusion;

c) relationships that organise the attachment of the national vocabulary to its reference thesaurus : hierarchy and compulsory multi-equivalence (in our sense of the term);

d) relationships that organise the different language versions of a multilingual national vocabulary : linguistic equivalence and "language adherence";

e) relationships that allow links between different national vocabularies : we shall introduce the relationships of inter- and intralinguistic "transnational equivalence".

a) relationships that organise "national concepts"

A semantic equivalence relationship links each non-preferred national term to one national term.

The linguistic equivalence relationship links the national terms which express the same concept in the different language versions in the case of a multilingual national vocabulary.

By analogy with the reference-thesaurus concepts, a **national concept** (CN) is made up of a set of these elements and relationships.

b) relationships that organise structuring

Hierarchical and associative relationships may be established between national concepts provided that the rules defined in the reference thesaurus are respected.

"National hierarchical chains" (NHC) may thus be developed.

As there are no concepts common to the national vocabularies and the reference thesaurus, associative relationships may only be established between national concepts.

Inclusive relationships may possibly be used to create subsystems analogous to microthesauri, i.e. **micro-national vocabularies (mNV),** in order to allow distribution of the national vocabulary into areas of meaning.

c) relationships that organise attachment

Ideally, to make re-indexing possible in all cases, each national concept must be linked, directly or indirectly, to one or several reference-thesaurus concept(s).

Cases where attachment is impossible have nevertheless been provided for. This situation must be regarded as temporary and should start off a procedure to draw the attention of the managers of the reference thesaurus to the possible gap that caused the problem.

Apart from this case, there are two methods of attachment : the "simple" hierarchical relationship and the compulsory multi-equivalence relationship (in the specific sense which we have given to it [114]).

Below, we mention the additional rules governing the hierarchical relationship which were proposed for attaching the sub-thesauri; we shall examine how far they may be applied to national vocabularies [115].

- Rule 1 is not applicable since there are no common concepts.

- Rule 2 is applicable : in the case of (a) hierarchical relationships or (b) compulsory multi-equivalence relationships established between national concepts and reference-thesaurus concepts,

  (a) the national concept is the specific and the reference-thesaurus concept is the generic;

  (b) the national concept refers to several reference-thesaurus concepts, not the other way round.

  It is worth noting that the compulsory multi-equivalence relationship is not allowed between national terms.

- Rule 3 is adapted and relaxed somewhat in accordance with what we have already stated. Ideally, each national concept is linked, either directly or indirectly, to at least one reference-thesaurus concept.

- Rule 4 (which advises against polyhierarchies) becomes useless insofar as the use of the compulsory multi-equivalence relationship, which is advised, will result in the re-indexings (see comment relating to rule 4B) leading to information-bearing combinations of national concepts.

  Duplication of the corresponding hierarchical chains in the national vocabulary (see comment relating to rule 4A) will not occur as there are no common concepts and the reference-thesaurus concepts which are "anchoring points" will not appear as entries in the structured presentation of the national vocabulary.

---

[114] See above, 5.2.1. b), page 65.

[115] See above, 3.2.3. a), page 37.

d) relationships that organise language versions

The linguistic equivalence and "language adherence" relationships are used to establish the different language versions of the national vocabularies corresponding to educational systems with more than one official language.

The official nature of multilingualism will erase the difficulty of establishing language equivalents, to which we alluded when dealing with the EUDISED Thesaurus[116]. The problem is now establishing "trans-national equivalence" relationships, which we shall introduce below.

e) relationships between national vocabularies

Some concepts are common to several educational systems.

If they are common to all the educational systems (possibly with one exception), their "vocation" is to become reference-thesaurus concepts[117].

If they are not, recognising and marking by an appropriate relationship the equivalences between concepts that belong to different national vocabularies lies within the scope of the third aim. This aim comes under coordination of the methodology for developing national vocabularies : these vocabularies must constitute "coordinated terminologies" which can be exploited with a view to locating the partial convergences between the different educational systems in Europe.

We propose a new type of relationship : the **trans-national** equivalence relationship between concepts that belong to different national vocabularies.

The following cases may be distinguished :

- the **inter-linguistic** *transnational equivalence* relationship between national concepts expressed in different languages (for instance, concepts common to the French and Italian educational systems);

- the **intra-linguistic** *transnational equivalence* relationship between national concepts expressed in the same language (for instance, concepts common to the Dutch-speaking educational systems in Belgium and the Netherlands).

The establishment of these transnational equivalence relationships leads to **transnational concepts (CTN)** being highlighted.

Such a concept is always *more* transnational *or less* : a concept may be common to two or more educational systems (should we say "transalpine concept" in the case of a concept common to the French and Italian educational systems ?) and to two or more language areas, etc.

This brings us back to the continuum presented above[118] . In this sense, the reference-thesaurus concept (C) corresponds to a particular case of transnational concept.

We shall show[119] that the process of establishing these relationships may be helped by locating :
- homographic national terms;
- concepts expressed in the same manner in the "common language".

---

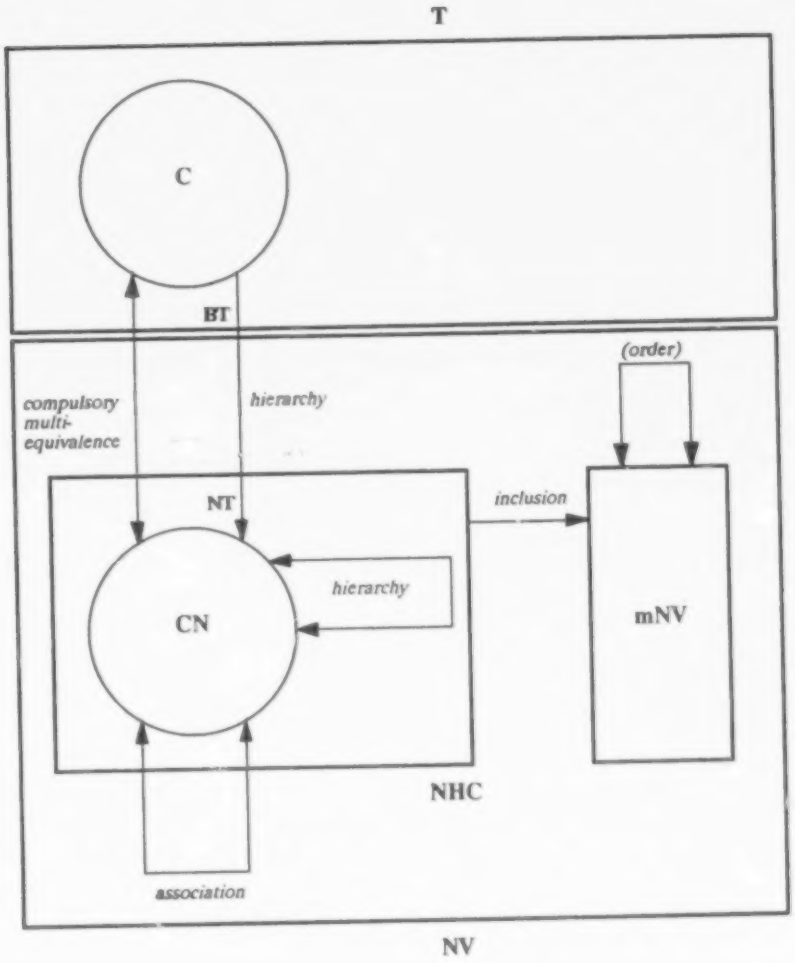
[116] See above, 2.2.3. f), page 22.

[117] See above, 5.2.1. b), page 65.

[118] See above, 5.2.1. a), page 63.

[119] See below, 5.2.5. b), page 76.

Figure 12. The national vocabulary as a system



5.2.4. Presentations

The initiative of publishing national vocabularies could be left to the bodies that develop them as this should entail no "automatic" feedback to the reference thesaurus.

Indeed, the proposals for introducing new descriptors and non-descriptors, which the development of the national vocabularies could suggest, will go through the usual procedures in the updating process of the European Education Thesaurus' content.

It will be to the advantage of the Metathesaurus promoters to offer technical support (software tool, for instance) to encourage these developments and to ensure dissemination which merits the effort that went into them.

It is also worth noting that taking the transnational equivalences into account requires their centralisation in a data management system to be implemented at community level[120].

Depending on the lessons of concrete experience, we think that it will be possible to present the national vocabularies in at least two of the three forms of presentation of the reference thesaurus, namely, structured alphabetical presentation and rotated alphabetical presentation, by language version if appropriate.

We shall not examine the conventional symbols which will have to be introduced to represent the compulsory multi-equivalence relationship (see Figure 10, page 67) and the transnational equivalence relationship.

5.2.5. Use

a) documentary use

Both the use of national vocabularies in the framework of documentary systems processing national information and the possibility of re-indexing indexations and search statements which were initially created using national terms in order to re-express them with reference-thesaurus descriptors are, by their principle, largely comparable to the case of sub-thesauri : same use, same operating methods.

All the consequences of what distinguishes national vocabularies from sub-thesauri must nonetheless be examined, namely, the fact that attachment by compulsory multi-equivalence relationships is used.

We shall examine successively the re-indexing operation and its consequences for information retrieval and the organisation of indexing.

- ***re-indexing***

  Applying the re-indexing operation to a national concept attached to the reference thesaurus by a compulsory multi-equivalence relationship results in a multiplication of indexing terms. For one indexing term in the initial indexation, there will be two or more reference-thesaurus descriptors in the resulting indexation.

  In addition to the inevitable depletion of information caused by any re-indexing in the specific-to-generic direction, there will be increased noise during retrieval if a larger number of indexing terms are merely juxtaposed.

---

[120] The corresponding data processing developments could aim to provide a decentralised tool whose data could be communicated to the central system and *vice versa*.

We shall refer to this method as "poor" re-indexing (it involves eliminating doubles).

Attempts should in fact be made to conserve the accuracy of the original indexing in the "common-language" indexing.

To this end, the combinations expressed by the compulsory multi-equivalence relationship must be respected so that the re-indexing operation restores, where necessary, groups of descriptors instead of a mere list, without eliminating doubles; we shall refer to this case as "rich" re-indexing.

Such enrichment of the re-indexing principle not only requires automation of the operation but also a corresponding adjustment of :

- the organisation of indexing,

- the tools used in the retrieval operation.

- *retrieval*

Indeed, the data processing system used for retrieving records which were indexed on the basis of a "rich" re-indexing must, in order to exploit this very richness, allow a distinction between :

- the occurrences of a descriptor which correspond to the presence of this descriptor in the original indexation;

- the occurences of the same descriptor when its presence is the result of a re-indexing carried out by means of a specific-to-generic hierarchical relationship (a national term has been replaced by a single descriptor);

- the occurences of this descriptor when its presence is the result of a re-indexing carried out by means of a compulsory multi-equivalence relationship (a national term has been replaced by a group of combined descriptors).

In the last case, the data processing system must be able to restore the various *groups* of descriptors separately so that they may be taken into account during retrieval.

Under these conditions, the accuracy of the original indexing -reinforced by the use of a national vocabulary- is totally exploited. As a consequence, retrievals can be more efficient if they are adequately formulated (search statements must mention the possible groups of descriptors to be searched).

- ***organisation of indexing***

The corresponding organisation of indexing must take account of the different systems available to those who want to share information.

The following must be stored separately :

- the (manual) original indexation using national terms ("indexation 1");

- the (manual) original indexation using reference-thesaurus descriptors ("indexation 2");

- the result of "poor" re-indexation (either automatic or semi-automatic) of "indexation 1" ("indexation3");

- the result of (automatic) "rich" re-indexing of "indexation 1" ("indexation 4").

This must enable the transfer of records containing the useful indexations into documentary systems which may be classified according to the following standpoints :

- the indexing language(s) which they use (the same national vocabulary and/or the reference thesaurus),

- the retrieval tool (which may or may not allow "rich" re-indexings to be exploited).

The situations corresponding to the various possible combinations of these two criteria can be represented in a double entry table which lists :

- in columns, the two possible situations as regards the indexing language(s) used :

  * joint use of the reference thesaurus and of the (same) national vocabulary : TR + VN,
  * use of the reference thesaurus only : TR;

- in lines, the two possible situations as regards the available retrieval tool, described as :

  * "sophisticated" if it allows "rich" re-indexing,
  * "plain" if it does not.

The squares in the table feature the serial number of the indexations which can be exploited and which it is therefore worthwhile transferring :

	TR + VN	TR
"sophisticated"	1 , 2	2 , 4
"plain"	1 , 2	2 , 3

"Indexation 2" (in common language) is transferred in all cases.

"Indexation 1" (in national terms) will be added to it for documentary systems using the same national vocabulary.

◆

For the systems which do not use a national vocabulary (or which do not use the same one), the result of a re-indexing will be added, i.e. :

- "indexation 4" ("rich") if they have the corresponding retrieval means,

- "indexation 3" ("poor") if they do not (it should be noted that joint storage of indexations 2 and 3 implies elimination of doubles).

b) extra-documentary use

The aim in this case is to emphasise the converging elements of the various European educational systems with a view to better mutual understanding, a condition for the desired approximation of these educational systems.

Remember that the exploitation of the national vocabularies as genuine "coordinated terminologies" depends on whether their compilers have been sufficiently rigorous and consistent in their approach to ensure comparable results.

Such exploitation is based on the existence of transnational concepts.

We shall deal successively with questions relating to :

- ways to allow the establishment of transnational equivalence relationships which form the basis of transnational concepts;

- possible feedback to the reference thesaurus of work exploiting the principle of the compulsory multi-equivalence relationship.

- **establishment of transnational equivalence relationships**

Transnational equivalences must be established by a group of experts in the field of comparative education, who can rely on experts in the various educational systems.

Their task will consist in formulating and checking hypotheses.

Hypotheses of equivalence can be made with a data processing system in which data relating to all the available national vocabularies have been integrated in order to locate :

- the national concepts expressed, in the same language, by homographic or quasi-homographic national terms;

- the national concepts linked by a hierarchical relationship to the same reference-thesaurus concept;

- the national concepts linked by a compulsory multi-equivalence relationship to the same combination of reference-thesaurus concepts.

Once these national concepts have been approximated, examining their respective definitions and consulting any source of information that may prove useful should allow confirmation or invalidation of the hypothesis which states that the underlying factors are close, and enable the extent to which they may be considered as "equivalent" to be determined.

This is because transnational equivalence -just like linguistic equivalence in the reference thesaurus- will always be relative; this is highlighted in the upside down approach below.

- **feedback to the reference thesaurus**

Indeed, another way to tackle the same problem is to start from the national terms which are known beforehand to express close factors, though they may not be regarded as completely "equivalent".

We think here of terms expressing realities whose "degree of equivalence" is precisely one of the concerns of those trying to find ways to :

- assess (and strengthen) the degree of convergence of the various European educational systems;

- encourage recognition of qualifications between the Member States to ensure freedom of movement, of establishment and of practice of a profession.

For instance, starting from terms such as the French BACCALAUREAT, the German ABITUR, the Italian MATURITA, etc., the following operations should be performed :

1- Expressing each of the corresponding national concepts by a combination of descriptors from the European Education Thesaurus. This operation should result in a series of partially similar "statements".

2- Checking whether the -real- differences which exist between these concepts may not be completely expressed by using a few different descriptors.

3- Establishing the criteria which allow these real differences to be expressed correctly and completely.

4- Examining whether these "differentiation criteria" are not due to the fact that some aspects of education are insufficiently or badly covered by the vocabulary in the European Education Thesaurus.

5- If this is the case, suggesting supplements to the thesaurus to provide better coverage of these aspects.

We have emphasised the positive feedback to the management of the European Education Thesaurus which could result.

The principle behind the method may be applied without any such concern. The establishment of grids of "equivalence/differentiation" criteria may be used to determine the degree of equivalence between qualifications, certification, levels of education, educational institutions, etc.

5.3. IMPLEMENTATION

It is clear that our exploration of the work frame which should be put to all those who wish to compile and systemise their national terminologies has led us to devise a system more complex than the sub-thesaurus.

This is largely due to the ambitious aims given to it. In particular, the will to turn the national vocabularies into a source of information which can be exploited outside the strictly documentary field has led us to favour the use of the compulsory multi-equivalence relationship as a method of attachment and it has prompted the interest of a transnational equivalence relationship.

There are important consequences as regards implementation.

a) We have briefly described the new functions of the software tools that should be available. Although there is little doubt that such data processing developments are technically feasible, they will still have to be decided on and created.

   From this angle, gathering under the single conceptual roof of the Metathesaurus all the projects whose development is being considered with a view to turning the European Education Thesaurus into an instrument which can adapt itself to a complex and changing environment offers the advantage of designing data processing tools based on an overall analysis of the needs and enabling medium-term planning of the necessary developments.

b) The various national parties will have to give total backing to these new designs.

   The experience of international cooperation acquired through managing the European Education Thesaurus is encouraging insofar as it was realised that the complexity of an instrument is no obstacle if the latter can give appropriate answers to all needs.

   In the case of national vocabularies, national differences are so great as regards information and documentation management in education that these needs are perceived differently.

   If the importance of what is at stake at Community level is simultaneously taken into account, it becomes clear that the necessary methodological coordination can only be made effective by a supra-national initiative, whose first task will be to have the proposed design ratified by all those involved.

c) It should be repeated that work relating to "extra-documentary" use[121] may only be based on existing national vocabularies. Some time will elapse between the moment when it is decided to develop national vocabularies in a coordinated way and the moment when they can be supplemented with transnational equivalence relationships, circulated outside their country of origin, etc.

That is why we recommend establishing the procedures (setting up groups of experts in comparative education) and perfecting the work method in pilot schemes which could rely on developments already made.

The complexity of the system described leads us to the idea of perhaps limiting the ambitions of the project to facilitate its construction.

In the national vocabularies, the compulsory multi-equivalence relationship was not resorted to as a method of attachment. This does not mean that such a relationship would not have been practical for the compilers; the fact is that they have not used this means whose possible application has not so far been mentioned.

The national vocabularies could nevertheless be built on a model not so different from that of sub-thesauri (though there would be no common portion with the reference thesaurus). The linking principle would here be that each concept which is a top term in the national vocabulary (where hierarchical chains may be developed) has a reference-thesaurus concept as its generic term.

The Austrian national vocabulary conforms to this design.

The model can be made even simpler -but then also poorer- if the possibilityof developing hierarchical chains specific to the national vocabulary is not provided for. In the French national vocabulary, for instance, each national concept is attached to a generic reference-thesaurus concept but there are practically no hierarchical relationships between national terms (they are only linked to one another by associative relationships).

In this case, the closest model is that of specialised vocabularies [122].

---

[121] See above, 5.2.5. b), page 76.

[122] See above, 3.4., page 46 and sqq.

CHAPTER 6 : TOWARDS THE EUROPEAN EDUCATION METATHESAURUS

The European Education Metathesaurus is not a thesaurus.

It is not a thick tome that could be found, from Athens to Dublin, on the desk of European documentalists specialised in education, but it could, one day, be the telematic "Sesame! ... " which gives them easy access to documentary information collected by most of their colleagues.

The Metathesaurus is the set made up of the European Education Thesaurus and the various indexing languages that will be attached to it, whether sub-thesauri, specialised vocabularies, co-thesauri or national vocabularies.

It is indeed a project insofar as it will only start existing when the first links with the "satellite" indexing languages are operational.

But it is also a project in essence because it is designed to develop constantly :

- its core, the European Education Thesaurus, will constantly transform itself, not only to keep up with the developments in the field which it covers, but also to accommodate the other indexing languages, firstly by providing them with satisfactory anchoring points and later by facilitating the feedback phenomenon, whose unavoidable yet productive nature has been highlighted;

- the number of sub-thesauri, specialised vocabularies, co-thesauri and national vocabularies will increase as time goes by;

- it may become necessary to devise other forms of indexing language to meet new kinds of needs which means appropriate methods of attachment will have to be designed.

This is because the ambition of the Metathesaurus project is to offer a conceptual and methodological framework sufficiently flexible to resolve a wide range of problems.

The Metathesaurus is the organisation of operational links between documentary languages. It may be viewed as a normative framework in which various parties are invited to work and which yields an essential benefit : the possibility of transnational and transdisciplinary communication which does not cause any reduction in the effectiveness of their own methods to index and retrieve documents.

To promote adherence to this "standard", the managers of the "European Education Metathesaurus" project should, in practice, facilitate as far as possible the desired coordination by making the necessary tools available, proposing a support for management, etc.

The ultimate aim of the "European Education Metathesaurus" project is to offer a "common language" to all the parties involved in European cooperation in education -in the broadest sense of the term.

To face up to the problems posed by its huge, changing, transdisciplinary nature while overcoming the obstacles of multilingualism and internationality which, after all, add spice to the European project, it became necessary to provide a technically consistent answer, based on sometimes long expressed needs, and to come up with the developments now achieved.

What is at stake is important. The increasingly wide and ambitious purposes which form the context of the project may be articulated as follows :

- the creation of conceptual tools aims to perfect operational management methods for the documentary languages;

- documentary languages are tools for information exchange;

- information exchange in education is a privileged means to improve mutual understanding;

- such improved understanding is the condition for approximating the different European educational systems;

- such approximation is one of the most vital elements in the construction of a People's Europe.

The European Education Metathesaurus is indeed in keeping with this. As a long-term project, it will form the reference context for all future developments.

ANNEX 1

LIST OF THE EUDISED THESAURUS MICROTHESAURI

LIST OF TERMINOGRAPHS

01 TEACHING AND TRAINING
02 LEARNING
03 PRINCIPLES OF EDUCATION
04 SYSTEM OF EDUCATION
05 ADMINISTRATION OF EDUCATION
06 EVALUATION
07 BUILDING
08 EQUIPMENT
09 CURRICULUM
10 CONTENT OF EDUCATION
11 SCIENCES AND TECHNOLOGY
12 SOCIAL SCIENCES
13 LANGUAGE SCIENCES
14 PHILOSOPHY AND RELIGION
15 ART
16 RESEARCH
17 INFORMATION SYSTEMS
18 INFORMATION SOURCE
19 PERSONALITY DEVELOPMENT
20 COGNITIVE PROCESS
21 PERSONALITY
22 AFFECTIVITY AND FEELING
23 BEHAVIOUR INCENTIVE
24 BEHAVIOUR
25 HEALTH
26 HANDICAP
27 INTERRELATIONS
28 GROUPS AND ORGANIZATIONS
29 SOCIO-CULTURAL ENVIRONMENT
30 DEMOGRAPHIC ENVIRONMENT
31 FAMILY ENVIRONMENT
32 ECONOMIC ENVIRONMENT
33 LABOUR ENVIRONMENT
34 PROFESSION AND PERSONNEL
35 ADMINISTRATION
36 PUBLIC ADMINISTRATION
37 INTERNATIONAL ORGANIZATIONS
38 AFRICA
39 AMERICA
40 ASIA/MIDDLE EAST/OCEANIA
41 EUROPE
42 REGIONS

ANNEX 2

PRESENTATION OF THE EUDISED THESAURUS

METHODS OF PRESENTATION

The EUDISED thesaurus is contained in one volume per language version. Each volume has three parts, in addition to the preface and introduction :

- a structured alphabetical presentation
- terminographs for each microthesaurus
- a rotated alphabetical presentation.

The first line at the top of each page is reserved for identification :

- top left corner : month and year of the thesaurus
- top right corner : in order,
  - . the EUDISED code : EU, the same for all language versions
  - . the language indicator specific to the language version
  - . the page number.

STRUCTURED ALPHABETICAL PRESENTATION

Structured alphabetical presentation, with three columns per page, includes two types of "entries" or "information blocks" :

- descriptor
- non-descriptor

listed alphabetically according to the complete designation of the descriptors and non-descriptors.

* Descriptor entry

- wording of the descriptor

- number(s) of the microthesaurus or microthesauri which the descriptor belongs

- language equivalents in each of the 8 o languages; each equivalent is preceded by language indicator; the language equivalents sorted in alphabetical order - constant from language to another - of language indicators

- scope note, where the descriptor has one

- non-descriptor(s) relevant to the entry descri and preceded by the symbol UF (USE FOR)

- generic descriptor(s) of the entry descri preceded by the symbol BT (BROADER TERM) an figure showing the number of hierarchical le between generic descriptor and entry descrip generic descriptors are listed :

  . firstly in order of ascending hierarch levels

  . within each hierarchical level in al betical order

BEST COPY AVAILABLE

- specific descriptor(s) of the entry descriptor preceded by the symbol NT (NARROWER TERM) and a figure showing the number of hierarchical levels; specific descriptors are listed :

  . firstly in order of descending hierarchical levels

  . within each hierarchical level in alphabetical order

- descriptor(s) associated with the entry descriptor preceded by the symbol RT (RELATED TERM) and listed alphabetically.

Example 12

```
SPECIAL EDUCATION
 (03
  D/SONDERSCHULWESEN
  F/ENSEIGNEMENT SPECIAL
  G/EIDIKH AGWGH
  I/EDUCAZIONE SPECIALE
  K/SPECIALUNDERVISNING
  N/BUITENGEWOON ONDERWIJS
  P/EDUCACAO ESPECIAL
  S/EDUCACION ESPECIAL
  SPECIAL TYPES OF EDUCATION FOR
  EXCEPTIONAL CHILDREN, MAINLY
  MENTALLY OR PHYSICALLY
   HANDICAPPED )
    UF   THERAPEUTIC EDUCATION
    BTI  EDUCATIONAL POLICY
    NTI  COMPENSATORY EDUCATION
    RT   EXCEPTIONAL
    RT   HANDICAPPED
    RT   SPECIAL SCHOOL
    RT   SPECIAL SCHOOL TEACHER
```

* Non-descriptor entry

- wording of the non-descriptor
- corresponding descriptor, preceded by the symbol USE.

Example 13

```
THERAPEUTIC EDUCATION
 USE  SPECIAL EDUCATION
```

TERMINOGRAPHS

Each of the 42 microthesauri is represented by a terminograph set out on a single page which is :

- identified in the bottom right corner by a two-digit serial number, identical in all language versions, followed by the title of the microthesaurus in the version in question (1)

- listed according to microthesaurus serial number.

Each terminograph contains an inner frame measuring about 15 x 20 cm (2) and containing all the descriptors in the microthesaurus. These descriptors are listed hierarchically under a top term (underlined) inside boxes (3).

Hierarchical relations between descriptors in the same microthesaurus are shown by offsetting specific descriptors to the right of the generic descriptors (4).

Relations of association between descriptors in the same microthesaurus are shown by (5) :

- thick lines for relations between top terms

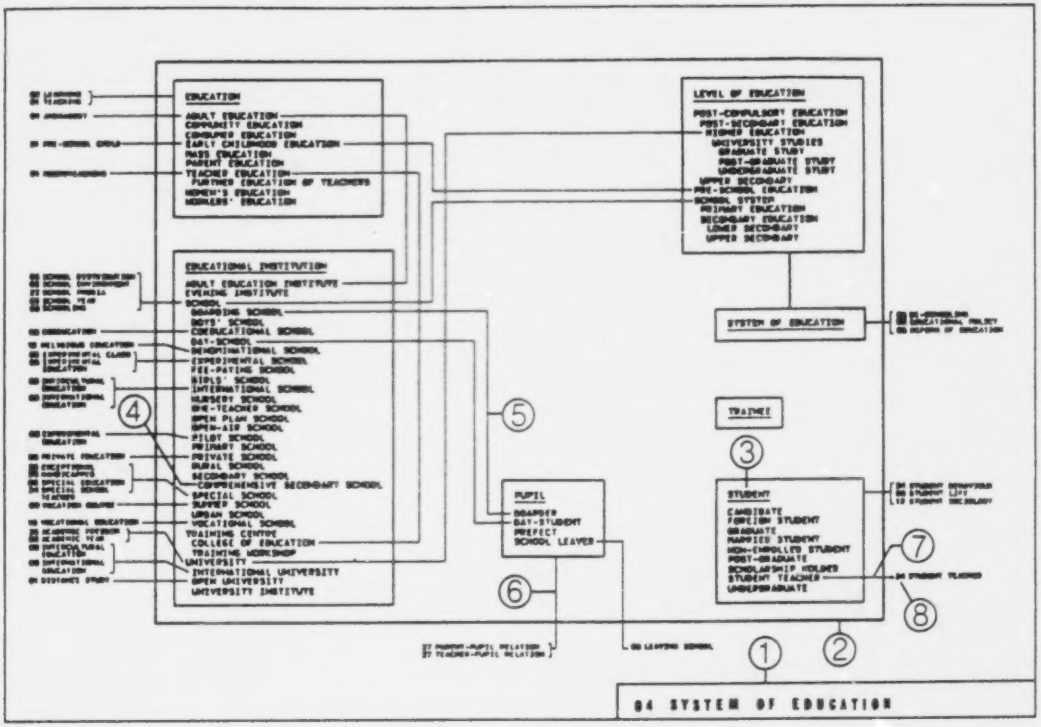
- medium lines for relations between a top term and another descriptor

- thin lines for relations between descriptors that are not top terms.

Outside the frame (2) appear descriptors referring to other microthesauri, but linked to the descriptors within the microthesaurus by some relation of association (shown by a line without an arrow) (6) or by a polyhierarchical relation (shown by a line with a double arrowed line) (7). On the left of each of these external descriptors appear(s) the number(s) of the microthesaurus to which it belongs. (8)

The non-descriptors are not marked on the terminographs.

BEST COPY A



Example 14

BEST COPY AVAILABLE

ROTATED ALPHABETICAL PRESENTATION

The rotated alphabetical presentation, with two columns per page, comprises two types of "entries" or "information blocks" :

- descriptor

- non-descriptor

alphabetically selected from each significant word in the descriptors and non-descriptors.

The information given under every such entry is as follows:

* descriptor entry

  - wording of the descriptor

  - number(s) of the microthesaurus or microthesauri containing the descriptor

* non-descriptor entry

  - wording of the non-descriptor

  - corresponding descriptor preceded by the symbol USE and followed by the number(s) of the microthesaurus or microthesauri containing it.

Example 15

THERAPEUTIC EDUCATION
USE: SPECIAL EDUCATION . . . 03
TRADITIONAL EDUCATION 03
VOCATIONAL EDUCATION 10
WOMEN'S EDUCATION 04
WORKERS' EDUCATION 04
EDUCATIONAL ADMINISTRATION
USE: ADMINISTRATION OF EDUCATION 05
EDUCATIONAL AIMS
USE: AIMS OF EDUCATION . . . 03
EDUCATIONAL BROADCASTING . . . 17
EDUCATIONAL BUILDING 07
EDUCATIONAL CHOICE
USE: CHOICE OF STUDIES . . . 10
EDUCATIONAL CRISIS
USE: CRISIS OF EDUCATION . . 03
EDUCATIONAL DEFICIT 03

BEST COPY AVAILABLE

ANNEX 3

INSTRUCTIONS TO THE EUDISED THESAURUS USER

INSTRUCTIONS FOR THE INDEXING OF DOCUMENTS

Indexing is the operation whereby the concepts dealt with in a document are identified and represented by descriptors in the thesaurus.

Its purpose is to make it possible to extract from all the bibliographical material either on index cards or on computer those which will give answers to the questions put; such extraction is possible only if the concepts in the question are formulated in the same language - that of the thesaurus descriptors.

The selection of concepts to be represented is based on two rules :

- selectivity : only those concepts for which the document supplies information likely to interest users must be selected

- exhaustiveness : all useful concepts contained in the text, whether explicit or implicit, must be selected.

BEST COPY AVAILABLE

Representation of the concepts thus selected, by means o
thesaurus descriptors, entails applying two other rules :

- vertical specificity : the descriptor must be on th
same level of specificity as the concept or, failin
this, on the level immediately above it in th
thesaurus

- horizontal specificity : a concept consisting of tw
or more words must be expressed by means of
composite descriptor, if there is one, rather tha
by a combination of single descriptors.

In concrete terms, representation of concepts
descriptors proceeds as follows :

- if the document is in a foreign language, th
concepts are translated into the language version
the thesaurus used by the indexer

- "entries" corresponding to the concepts expressed
the document are looked for in the rotate
alphabetical presentation of the thesaurus. The
are three possibilities :

  . expression of the concept corresponds to
  descriptor signifying it : this is entered
  the indexing form

  . expression of the concept corresponds to
  non-descriptor signifying it : the latte
  refers to the descriptor to be used; the
  proceed as above

. there is no entry corresponding to the expression of the concept. Two methods may then be used :

* either think of another formulation of the concept in natural language and see whether it corresponds to an entry in the thesaurus

* or define the general class or classes (micro-thesaurus) covering the recalcitrant concept, consult the terminographs and look among available descriptors for those which best represent the concept.

Correct application of the rules of selectivity, exhaustiveness and specificity is essential to the quality of a documentary information storage and retrieval system.

- Selectivity of indexing :

. is independent of the quality of the thesaurus

. depends on the indexing policy applied for each documentary system

. depends on the indexing documentalist's knowledge of the users and their needs.

- Exhaustiveness is conditioned :

. by the wealth of content of the thesaurus; obviously only those concepts which are present in the documentary language can be indexed

. by indexing policy, which establishes depth
indexing : it is not possible to determine th
depth in advance; there is an optimum whi
depends on the nature of the questions a
which can only be determined iteratively

. by the behaviour of the documentalists (in t
absence of a specific indexing policy) a
their meticulousness.

An increase in the indexing exhaustivene
influences information retrieval in two ways :

. it yields an improvement in response, i.e. t
percentage of relevant documents extracted fr
the thesaurus in reply to a question

. it causes a drop in precision, i.e. t
proportion of truly relevant documen
extracted in reply to a question.

If all concepts, even very subsidiary ones, prese
in the documents are indexed, the chances
retrieving all or almost all the relevant documen
in reply to the questions put are increased, but
the same time the bibliographies obtained wi
contain a high percentage of documents which are
little or no relevance.

- Specificity of indexing also depends on :

. the thesaurus, which may or may not include
large number of specific descriptors

. the indexing policy

. the documentalists.

Unlike increased exhaustiveness, an increase in specificity yields :

. a drop in the number of documents retrieved

. an improvement in precision.

If documents are indexed by means of descriptors having the same level of specificity as the concepts which they contain, questions concerning these descriptors will produce a bibliography that will not include those documents which are indexed at a more generic level, some of which in fact will contain relevant information; by way of compensation, however, the list will include few irrelevant documents. It should be noted, however, that documents indexed at the more generic level may be retrieved precisely through a research among these descriptors.

The reverse is unfortunately not the case. If indexing is not specific we shall have :

. better retrieval

. poorer precision.

For that reason, it is important to be specific when indexing documents.

INSTRUCTIONS FOR THE FORMULATION OF QUESTIONS

The formulation of a question and indexing are symmetrical operations : indexing consisted of identifying the concepts dealt with in a document on which information likely to be of interest to users was given; these concepts were then designated by descriptors. Formulation of a question consists of identifying the descriptors which may have served to translate the concepts characterising information that satisfied a demand and subsequently extracting the indexed documents by means of the descriptors.

Identification of the concepts in a question calls for a very different approach from that used in identifying the concepts in a document : the indexing of a document is an operation that leads to a high degree of compression - a document of 5 to 10 pages is generally represented by 5 to 15 descriptors (i.e. a compression rate of between 1% and 1 per mil) while the formulation of a question usually leads to expansion of the initial wording.

The first stage in the formulation of a question is to make explicit in as much detail as possible the user's information requirement, expressed in natural language - the documentalist identifies the concepts which together comprise the wording of the user's question.

The documentalist must then represent these concepts by means of the thesaurus descriptors. Three cases should be distinguished :

- the user's question is precise and the concepts contained in it are present in the thesaurus : the documentalist will preferably make use of the complete structured alphabetical list in which he will find :

  . the descriptors or non-descriptors (and thence other descriptors) corresponding to the wording of the concepts in the question

  . descriptors in hierarchical relations and relations of association which may help him to broaden the question

- not all or none of the concepts contained in the question appear in the thesaurus : after looking for them in vain in the structured alphabetical list or in the rotated alphabetical list, the documentalist will explore the terminographs corresponding to the field(s) covered by the question put and look in them for existing descriptors designating concepts nearest to those in the question

- the user's question is imprecise, as is frequently the case : the documentalist will preferably use terminographs, exploring the one(s) which correspond(s) to the user's field(s) of interest : there he will find an ordered collection of descriptors covering the field(s) in question and he will be able to choose the most suitable descriptors.

ANNEX 4

LIST OF THE CEDEFOP THESAURUS TOP TERMS

ARTS
(D/KUNST - F/ARTS - G/TEXNEI
I/ARTI - P/ARTES - S/ARTES)
NT1 CINEMA
NT1 DANCE
NT1 FINE ARTS
NT1 FOLK ARTS
NT1 HANDICRAFT
NT1 LITERATURE
NT1 MUSIC
NT1 THEATRE

BEHAVIOUR
(D/VERHALTEN - F/COMPORTEMENT
G/ΣΥΜΠΕΡΙΦΟΡΑ - I/COMPORTAMENTO
P/COMPORTAMENTO
S/COMPORTAMIENTO
FOR GENERAL ITEMS ON HUMAN
BEHAVIOUR)
NT1 BEHAVIOURAL CHANGE
NT1 HABITS
NT1 INDIVIDUAL BEHAVIOUR
NT1 SOCIAL BEHAVIOUR

COMMUNICATION
(D/KOMMUNIKATION - F/COMMUNICATION
G/ΕΠΙΚΟΙΝΩΝΙΑ - I/COMUNICAZIONE
P/COMUNICAÇAO - S/COMUNICACION)
NT1 COMMUNICATION MEDIA
NT1 COMMUNICATION SKILLS
NT1 COMMUNICATIONS INDUSTRY
NT1 INTERGROUP COMMUNICATION
NT1 INTERPERSONAL COMMUNICATION
NT1 NON-VERBAL COMMUNICATION
NT1 VERBAL COMMUNICATION

COUNTRIES AND REGIONS
(D/LAENDER UND GEBIETE
F/PAYS ET REGIONS
G/ΧΩΡΕΙ ΚΑΙ ΠΕΡΙΟΧΕΙΕΙ
I/PAESI E REGIONI
P/PAISES E REGIOES
S/PAISES Y REGIONES)
NT1 AFRICA
NT1 AMERICAS
NT1 ARAB COUNTRIES
NT1 ASIA
NT1 CAPITALIST COUNTRIES
NT1 DEVELOPED COUNTRIES
NT1 DEVELOPING COUNTRIES
NT1 EUROPE
NT1 FOREIGN COUNTRIES
NT1 MEDITERRANEAN COUNTRIES
NT1 MIDDLE EAST
NT1 OCEANIA
NT1 OECD COUNTRIES
NT1 SOCIALIST COUNTRIES

ECONOMY
(D/WIRTSCHAFT - F/ECONOMIE
G/OIKONOMIA - I/ECONOMIA
P/ECONOMIA - S/ECONOMIA)
NT1 BUSINESS ECONOMICS
NT1 CONSUMPTION

ECONOMY (C)
NT1 ECONOMIC COMPETITION
NT1 ECONOMIC CONCENTRATION
NT1 ECONOMIC CONDITIONS
NT1 ECONOMIC POLICY
NT1 ECONOMIC RESOURCES
NT1 ECONOMIC SECTORS
NT1 ECONOMIC SYSTEMS
NT1 INTERNATIONAL ECONOMY
NT1 MARKET
NT1 PRICES
NT1 PRODUCTION
NT1 PRODUCTIVITY
NT1 PUBLIC FINANCE

EDUCATION AND TRAINING
(D/BILDUNG UND AUSBILDUNG
F/EDUCATION ET FORMATION
G/ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΚΑΤΑΡΤΙΣΗ
I/ISTRUZIONE E FORMAZIONE PROFES.
P/EDUCAÇAO E FORMAÇAO
S/EDUCACION Y FORMACION)
NT1 EDUCATION
NT1 TRAINING

ENVIRONMENT
(D/UMWELT - F/ENVIRONNEMENT
G/ΠΕΡΙΒΑΛΛΟΝ - I/AMBIENTE
P/AMBIENTE - S/MEDIO AMBIENTE)
NT1 CLIMATE
NT1 ENVIRONMENTAL ENGINEERING
NT1 ENVIRONMENTAL PROTECTION
NT1 NATURAL RESOURCES
NT1 POLLUTION

HANDICAPPED PERSONS
(D/BEHINDERTE - F/HANDICAPES
G/ΜΕΙΟΝΕΚΤΟΥΝΤΑ ΑΤΟΜΑ
I/HANDICAPPATI - P/DEFICIENTES
S/MINUSVALIDOS)
NT1 HANDICAPPED WORKERS
NT1 MENTALLY HANDICAPPED PERSONS
NT1 PHYSICALLY HANDICAPPED PERSONS
NT1 SOCIALLY HANDICAPPED PERSONS

HEALTH
(D/GESUNDHEIT - F/SANTE - G/ΥΓΕΙΑ
I/SALUTE - P/SAUDE - S/SALUD)
NT1 DISEASES
NT1 HEALTH ADMINISTRATION
NT1 HEALTH EDUCATION
NT1 HEALTH POLICY
NT1 HEALTH SERVICES
NT1 MENTAL HEALTH
NT1 PHYSICAL DEVELOPMENT
NT1 PHYSICAL HEALTH

INFORMATION
(D/INFORMATION - F/INFORMATION
G/ΠΛΗΡΟΦΟΡΙΑ - I/INFORMAZIONE
P/INFORMAÇAO - S/INFORMACION
THE FACTS OR CONCEPTS THAT ARE
DISSEMINATED IN ORDER TO INCREASE
KNOWLEDGE)

INFORMATION (C)
NT1 ACCESS TO INFORMATION
NT1 DOCUMENTATION
NT1 INFORMATION DISSEMINATION
NT1 INFORMATION MATERIALS
NT1 INFORMATION TECHNOLOGY
NT1 INFORMATION USE
NT1 TRAINING INFORMATION
NT1 VOCATIONAL INFORMATION

INTERNATIONAL ORGANIZATIONS
(D/INTERNATIONALE ORGANISATIONEN
F/ORGANISATIONS INTERNATIONALES
G/ΔΙΕΘΝΕΙΣ ΟΡΓΑΝΙΣΜΟΙ
I/ORGANIZZAZIONI INTERNAZIONALI
P/ORGANIZAÇOES INTERNACIONAIS
S/ORGANIZACIONES INTERNACIONALES)
NT1 INTERGOVERNMENTAL ORGANIZATIONS
NT1 NON-GOVERNMENTAL ORGANIZATIONS

INTERNATIONAL RELATIONS
(D/INTERNATIONALE BEZIEHUNGEN
F/RELATIONS INTERNATIONALES
G/ΔΙΕΘΝΕΙΣ ΣΧΕΣΕΙΣ
I/RELAZIONI INTERNAZIONALI
P/RELAÇOES INTERNACIONAIS
S/RELACIONES INTERNACIONALES)
NT1 INTERNATIONAL AGREEMENTS
NT1 INTERNATIONAL COOPERATION

LABOUR
(D/ARBEIT - F/TRAVAIL - G/ΕΡΓΑΣΙΑ
I/LAVORO - P/TRABALHO - S/TRABAJO)
NT1 EMPLOYMENT
NT1 INDUSTRIAL PSYCHOLOGY
NT1 INDUSTRIAL SOCIOLOGY
NT1 LABOUR ADMINISTRATION
NT1 LABOUR ECONOMICS
NT1 LABOUR LEGISLATION
NT1 LABOUR MARKET
NT1 LABOUR POLICY
NT1 LABOUR RELATIONS
NT1 LABOUR STATISTICS
NT1 PERSONNEL
NT1 PERSONNEL MANAGEMENT

OCCUPATIONS
(D/BERUFE - F/PROFESSIONS
G/ΕΠΑΓΓΕΛΜΑΤΑ - I/PROFESSIONI
P/PROFISSOES - S/PROFESIONES)
NT1 ADMINISTRATIVE PERSONNEL
NT1 AGRICULTURAL WORKERS
NT1 CLERICAL WORKERS
NT1 CRAFTSMEN
NT1 INDUSTRIAL WORKERS
NT1 PROFESSIONAL WORKERS
NT1 SALES WORKERS
NT1 SERVICE WORKERS
NT1 TRANSPORT WORKERS

POLITICAL SYSTEMS
(D/POLITISCHE SYSTEME
F/REGIMES POLITIQUES
G/ΠΟΛΙΤΙΚΑ ΣΥΣΤΗΜΑΤΑ

1 - 1987

POLITICAL SYSTEMS (C)
I/SISTEMI POLITICI
P/SISTEMAS POLITICOS
S/SISTEMAS POLITICOS)
NT1 ADMINISTRATION
NT1 HUMAN RIGHTS
NT1 INSTITUTIONAL FRAMEWORK
NT1 POLITICS

POPULATION
(D/BEVOELKERUNG - F/POPULATION
G/ΠΛΗΘΥΣΜΟΣ - I/POPOLAZIONE
P/POPULAÇAO - S/POBLACION)
NT1 HUMAN RESOURCES
NT1 MARRIAGE
NT1 MIGRATION
NT1 POPULATION POLICY
NT1 POPULATION TRENDS
NT1 SEX
NT1 STUDENT POPULATION

RESEARCH
(D/FORSCHUNG - F/RECHERCHE
G/EPEYNA - I/RICERCA
P/INVESTIGAÇAO - S/INVESTIGACION)
NT1 METHODOLOGY
NT1 ORGANIZATION OF RESEARCH
NT1 THEORY
NT1 TRAINING RESEARCH

SCIENCE AND TECHNOLOGY
(D/WISSENSCHAFT UND TECHNOLOGIE
F/SCIENCE ET TECHNOLOGIE
G/ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ
I/SCIENZA E TECNOLOGIA
P/CIENCIA E TECNOLOGIA
S/CIENCIA Y TECNOLOGIA)
NT1 SCIENCES
NT1 TECHNOLOGY

SOCIETY
(D/GESELLSCHAFT - F/SOCIETE
G/ΚΟΙΝΩΝΙΑ - I/SOCIETA
P/SOCIEDADE - S/SOCIEDAD)
NT1 COMMUNITIES
NT1 SOCIAL ADMINISTRATION
NT1 SOCIAL CHANGE
NT1 SOCIAL CONTROL
NT1 SOCIAL ENVIRONMENT
NT1 SOCIAL GROUPS
NT1 SOCIAL MOVEMENTS
NT1 SOCIAL POLICY
NT1 SOCIAL PROBLEMS
NT1 SOCIAL STRUCTURE

ANNEX 5

ABBREVIATIONS

N.B. The asterisked terms appear in the lexicon in Annex 6.

BT	Broader term
C	Concept
CN	National concept
CST	Sub-thesaurus concept
CTN	Transnational concept
D	Descriptor
DST	Sub-thesaurus descriptor
HC	Hierarchical chain
LV	Language version
mNV	Micro national vocabulary
mT	Microthesaurus*
MT	Metathesaurus*
NDST	Sub-thesaurus non-descriptor
nT	Nanothesaurus*
NHC	National hierarchical chain
NT	Narrower term
NTN	Non-preferred national term
NV	National vocabulary*
RT	Related term
ST	Sub-thesaurus*
SV	Specialised vocabulary*
T	Thesaurus
TN	National term
TS	Specialised term
TT	Top term
UF	Used for
UFS	Used for from sub-thesaurus or from specialised vocabulary
USE	Use

ANNEX 6

LEXICON

N.B. *Italicised* terms are entries in this lexicon.

Associated thesauri

Expression used to designate a pair of thesauri having the features required for "association". Before and during of the approximating process, they will be referred to rather as *reference thesaurus* to which a *co-thesaurus* is being associated.

Co-thesaurus

Expression used to designate a thesaurus which should be associated to a given *reference thesaurus;* at the end of the approximating process, the two thesauri are called *associated thesauri*.

EUDISED Thesaurus

Multilingual thesaurus for the processing of information in the field of education, whose third edition is becoming the *European Education Thesaurus* .

European Education Thesaurus

The new name of *EUDISED Thesaurus*. It is destined to become the core of the *Metathesaurus*.

Metathesaurus (MT)

Expression used to designate the complex system encompassing the *European Education Thesaurus* and the various forms of indexing language linked to it : *sub-thesauri* (possibly *specialised vocabularies), co-thesauri* and *national vocabularies*.

Microthesaurus (mT)

Set of hierarchical chains of the *European Education Thesaurus* which correspond to a distribution into areas of meaning.

Nanothesaurus (nT)

The equivalent, for a *sub-thesaurus*, of a *microthesaurus*, i.e. a set of hierarchical chains corresponding to a distribution into areas of meaning.

National vocabulary (NV)

A form of semantically structured indexing language characterised by a thematic coverage corresponding to the specific realities of education in a given educational system. It is linked to a *reference thesaurus* by relationships between its "national terms" and the descriptors of the *reference thesaurus*.

Reference thesaurus

Expression used to designate the *European Education Thesaurus* when it is regarded as the core of the *Metathesaurus*, i.e. in its position as a "common language" which enables communication between users of the different forms of indexing language which are linked to it : namely, *sub-thesauri* (possibly *specialised vocabularies*), *co-thesauri*, and *national vocabularies*.

Specialised vocabulary (SV)

A form of non semantically structured indexing language characterised by in-depth thematic coverage of a restricted number of fields. It is not linked to its *reference thesaurus* by common concepts but only by relationships between "specialised terms" and descriptors of the *reference thesaurus*.

Sub-thesaurus (ST)

A form of semantically structured indexing language characterised by an in-depth thematic cover of a restricted number of fields. A sub-thesaurus is linked to its *reference thesaurus* by common concepts and hierarchical relationships governed by particular rules.

ANNEX 7

BIBLIOGRAPHY

CNDP;- "Liste des additifs au thésaurus EUDISED : Document de travail"; -not published, 1987, 64 p.

COUNCIL OF EUROPE;- "Vocabulaire destiné à l'indexage de documents en matière d'enseignement de langues vivantes en lien avec le thésaurus EUDISED", prepared by P.KNOPF; - Strasbourg : Council of Europe, 1987, 41 p., DECS/Doc (87) 24.

COUNCIL OF EUROPE;- "EUDISED Thesaurus : Appendix 1 : Austrian national descriptors", list prepared by the Austrian Ministry of Education, Culture and Sports;- Strasbourg : Council of Europe, 1989, 33p., DECS/Doc (89) 7.

DAHLBERG, I.;- "Conceptual compatibility of ordering systems" in Int. Classif.10 (1983) No.1, p. 5-8.

ISO;- "Guidelines for the establishment and development of multilingual thesauri";- Geneva : ISO, 1985, 61p., ISO 5964-1985 (E).

ISO;-" Guidelines for the establishment and development of monolingual thesauri";-Geneva : ISO, 1986, 32 p., ISO 2788-1986 (E).

UNESCO; "Unesco Education Thesaurus: IBE ";- Paris : UNESCO, 1979, 319 p.

VAN SLYPE, G.;- "Les langages d'indexation : conception, construction et utilisation dans les systèmes documentaires";- Paris : Les Editions d'Organisation, 1987, 277p.

VIET, J. and VAN SLYPE, G.; -"EUDISED, multilingual thesaurus for information processing in the field of education ";- Berlin : Mouton Publ., 1984, 302p.

VIET, J.;- "Thesaurus of vocational training - CEDEFOP "; Luxemburg : Opoce, 1988, 218 p.

WHITELOCK, P. J.;- "A descriptor bank of social science terms" in Int.Classif.9 (1982) No.3, p. 145-151.

Eurydice
The Education Information Network in the European Community

Educational cooperation in the Community

The educational systems in Community countries vary considerably and this variety, which is the result of historic and cultural factors, itself constitutes a great wealth which should be preserved.
On the other hand it is in the interest of everyone to draw benefit from the experiences and projects of partner countries within the Community.

Educational policy makers in the European Community have long been aware of the value of information concerning innovations and changes introduced in other Member States in order that their own decisions might be well founded on the knowledge of alternative possibilities of practice, and the Member States agreed to promote close cooperation in the field of education at Community level.

In February 1976 the Council of the European Communities and the Ministers of Education adopted an action programme in the field of education. They agreed among other things to set up an information Network in order to increase and improve the circulation of information in the area of education policy*.
This information Network, known as EURYDICE, is designed therefore to underpin the developing programme of educational cooperation within the European Community.

The Education Committee of the European Community which has the task of coordinating and monitoring the action programme recommended that the use of Eurydice should in the first phase be limited to those in policy making positions. The Education Committee consists of representatives of the Member States and of the Commission.

In accordance with their own distinctive education structures, each Member State has designated at least one National Unit to participate in the Network and the Commission of the European Communities has arranged with the European Cultural Foundation the establishment of the European Eurydice Unit, which is part of the Network.
Eurydice is based on the mutual cooperation of the national units and the European Unit. In addition the latter coordinates and animates the network.

* Official Journal n°C38, 19 [illegible]